Maastricht University

Autoregressive Wild Bootstrap Inference for Nonparametric Trends

Marina Friedrich, Stephan Smeekes, Jean-Pierre Urbain

Autoregressive Wild Bootstrap Inference for Nonparametric Trends

RM/17/010

GSBE

Maastricht University School of Business and Economics
Graduate School of Business and Economics

P.O Box 616
NL- 6200 MD Maastricht
The Netherlands

Autoregressive wild bootstrap inference for nonparametric trends

Marina Friedrich[*] Stephan Smeekes[*] Jean-Pierre Urbain[†]

Department of Quantitative Economics

Maastricht University

April 5, 2017

Abstract

In this paper a modified wild bootstrap method is presented to construct pointwise confidence intervals around a nonparametric deterministic trend model. We derive the asymptotic distribution of a nonparametric kernel estimator of the trend function under general conditions, which allow for serial correlation and heteroskedasticity. Asymptotic validity of the bootstrap method is established and it is shown to work well in finite samples in an extensive simulation study. The bootstrap method has the potential of providing simultaneous confidence bands for the same models along the lines of Bühlmann (1998) and can be applied without further adjustments to missing data. We illustrate this by applying the proposed method to a time series of atmospheric ethane which can be used as an indicator of atmospheric pollution and transport.

JEL classifications: C14, C22.

Keywords: autoregressive wild bootstrap, nonparametric estimation, time series, simultaneous confidence bands, trend estimation

[*]Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. E-mail: m.friedrich@maastrichtuniversity.nl, s.smeekes@maastrichtuniversity.nl. We would like to thank Eric Beutner and Hanno Reuvers for helpful discussions as well as Franz Palm and David Hendry for valuable comments. Previous versions of this paper were presented at the conference Econometric Models of Climate Change in Aarhus and at the Maastricht Workshop on Advances in Quantitative Economics II, as well as at a seminar at Oxford University. We thank all participants of the conference, the workshop and the seminar for their comments and suggestions. We further thank Whitney Bader, Bruno Franco, Bernard Lejeune and Emmanuel Mahieu from the Institute of Astrophysics and Geophysics, University of Liège, for the ethane data. These data are available from the Network for the Detection of Atmospheric Composition Change website at ftp://ftp.cpc.ncep.noaa.gov/ndacc/station/jungfrau/hdf/ftir/.

[†]Deceased on October 1, 2016

1 Introduction

Many variables of interest in economic and financial research are changing over time and trend estimation has become a crucial tool for any analysis involving these variables. Not only economic variables display a trending behavior; another interesting field of application of trend modeling is climatology. It is of major concern to researchers in this area to determine whether environmental time series exhibit an upward trend. While it is simple to capture trending behavior by inclusion of a linear trend, the true underlying trend is mostly of a more complex form. A miss-specified trend model is likely to be the result when estimating a linear trend under these circumstances. To take into account the complexity of the data, a flexible model is needed. A large body of econometric research therefore focuses on nonparametric modeling and estimation, which is known for its flexibility.

In particular, when modeling trends in temperature and emission data researchers find an increased need to rely on more advanced econometric tools due to correlation present in the data. This has been addressed in a parametric framework by, e.g., Franses and Vogelsang (2005) and McKitrick and Vogelsang (2014) who investigate various temperature series. Many economic and climatological time series displaying trending behavor also exhibit heteroskedasticity, making inference on these trends a challenging task. To address the problems of heteroskedastic and autocorrelated innovations we rely on bootstrap methods. In the presence of serial correlation, an autoregressive sieve or block bootstrap method can provide valid inference. For the nonparametric trend model this approach has been used by Bühlmann (1998) who shows the validity of a sieve bootstrap for general forms of dependence. In addition, Neumann (1997) uses a wild bootstrap method to achieve robustness to heteroskedasticity for a similar model. The wild bootstrap approach is also suitable with missing data, as advocated for example by Shao (2010) in a different context. In particular in climatology, this feature of the wild bootstrap offers an important benefit over other methods, since there is no need of imputing missing data points. They constitute a problem which is frequently encountered in this strand of research, which could be due to instrument failure or unfavorable measurement conditions. This topic has, to our knowledge, not yet been explicitly addressed in the nonparametric trend literature. The wild bootstrap, however, relies on independence of the error terms, which is a situation rarely encountered in practice. To relax this strong assumption, dependent versions of wild bootstrap methods have been proposed in other contexts - see Shao (2010), Leucht und Neumann (2013) and Smeekes and Urbain (2014) for examples of this - but not in the context of nonparametric trend estimation.

For nonparametric trend estimation simultaneous confidence bands are more informative than pointwise confidence intervals. Some research questions, like whether there has been an upward trend over a certain period of time, can be addressed with simultaneous confidence bands. Wu and Zhao (2007) derive such bands for the nonparametric trend model and these have asymptotically correct nominal coverage probabilities. Bühlmann (1998) proposes sieve bootstrap-based simultaneous confidence bands that are not only asymptotically valid but also have good small sample performance. They can, however, not easily be adjusted to be applicable to time series with missing data.

In this paper, we present a dependent version of the wild bootstrap method, the autoregressive wild bootstrap, to construct confidence intervals for the nonparametric trend model under general conditions which allow for serial correlation and heteroskedasticity. The bootstrap method has the potential of providing simultaneous confidence bands along the lines of Bühlmann (1998) and can be applied without further adjustments to missing data.

To illustrate our methodology, we study a time series of atmospheric ethane emissions for which almost 70% of the data points are missing. The series has previously been investigated by Franco et al. (2015). Atmospheric ethane is an indirect greenhouse gas which can be used as an indicator of atmospheric pollution and transport. It is emitted during shale gas extraction and since shale gas has become more and more important as a source of natural gas, geophysicists and climatologists are interested in analyzing trends in such data.

The paper is organized as follows. In Section 2, we set the ground by introducing the model. Section 3 describes the estimation procedure and how to obtain confidence intervals. It gives the bootstrap algorithm and explains how to construct simultaneous confidence bands. Subsequently, Section 4 presents the asymptotic distribution of the estimator and establishes asymptotic validity of the autoregressive wild bootstrap. In addition, the theoretical results for the construction of simultaneous confidence bands are derived. Section 5 presents results from a simulation study which tests the finite sample performance. We discuss an application to a time series of atmospheric ethane in Section 6, while Section 7 concludes. All proofs are given in the Appendix.

To conclude the introduction, we give some notation. The time series of interest $\{y_t\}_{t=1}^n$ is defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω denotes the sample space, $\mathcal{F}$ a sigma-algebra on Ω and $\mathbb{P}$ a probability measure. We will denote by $\xrightarrow{d}$ weak convergence and by $\xrightarrow{p}$ convergence in probability. Whenever a quantity has a subscript * it denotes a bootstrap quantity, conditional on the original sample. Bootstrap weak convergence in probability is denoted by $\xrightarrow{d^*}_p$. $\lfloor x \rfloor$ stands for the largest integer smaller than or equal to x.

2 A deterministic trend model

We consider the following deterministic trend model

$$y_t = m_t + \sigma_t u_t = m_t + z_t \qquad t = 1, ..., n \tag{2.1}$$

where t is a time index, $z_t = \sigma_t u_t$ and $m_t = m(t/n)$ is the deterministic trend with $m : [0, 1] \to \mathbb{R}$ being a smooth function (see Assumption 1). Similarly, $\sigma_t = \sigma(t/n)$ with the function $\sigma : [0, 1] \to \mathbb{R}^+$ captures heteroskedasticity as allowed for by Assumption 2. Finally we assume that $\{u_t\}_{t=1}^n$ is a linear process

$$u_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \quad \text{with} \quad \psi_0 = 1, \tag{2.2}$$

which satisfies Assumption 3 below.

Assumption 1. *$m(\cdot)$ is twice continuously differentiable deterministic function on $(0, 1)$ with $\sup_{0<\tau<1} |m^{(i)}(\tau)| < \infty$, for $i = 0, 1, 2$, where $m^{(i)}(\cdot)$, $i = 0, 1, 2$, stand for the function itself* *and its first and second derivative, respectively.*

Assumption 2. $\sigma(\cdot) : [0,1] \to \mathbb{R}^+$ *is a deterministic function that satisfies Lipschitz continuity.*

Assumption 3. *(i)* $\{\epsilon_t\}$ *is i.i.d with* $\mathbb{E}(\epsilon_t) = 0$, $\mathbb{E}(\epsilon_t^2) = 1$, *and* $\mathbb{E}(\epsilon_t^4) < \infty$. *(ii)* $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ *and the lag polynomial* $\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j \neq 0$ *for all* $z \in \mathbb{C}$ *and* $|z| \leq 1$.

The smoothness assumptions of the object to be estimated, $m(\cdot)$, are given in Assumption 1. This assumption is crucial for the estimation method to work and it can be seen as quite restrictive. Given the potential applications for our method to climatological time series, however, it might not be such a strong assumption to make, as climatological processes tend to be such that change occur gradually. In addition, many of these series are measured daily or even multiple times a day such gradual change would have to occur rather rapidly for it to appear like an abrupt break.

Assumption 2 allows for a wide array of unconditional heteroskedasticity. In its current form it excludes abrupt breaks, although these could be allowed by by generalizing the function $\sigma(\cdot)$ to be piecewise Lipschitz as in Smeekes and Urbain (2014). However, given that allowing for abrupt breaks is not the most relevant for our setup, we do not pursue this in the current paper.

Assumption 3 is a standard linear process assumption that ensures that sufficient moments of $\{u_t\}$ exist, that the linear process satisfies a summability condition and that the polynomial is invertible such that $\{u_t\}$ is strictly stationary. These assumptions are satisfied by large class of stationary processes including, but not limited to, all finite order stationary ARMA models. As our bootstrap method does not require linearity, alternative dependence concepts such as mixing, which was considered in the same bootstrap context by Smeekes and Urbain (2014), could be used as well; the advantage of the linear process setup is that it makes our setting directly comparable to Bühlmann (1998) except he does not consider unconditional heteroskedasticity.

3 Inference

The object of interest in this paper is the trend function $m(\cdot)$ defined in Section 2 which we estimate given the observations $y_1, ..., y_n$. In this section we give a detailed description of how to perform inference on the model introduced in the previous section. We first describe the nonparametric trend estimation method, and treat the bootstrap method next.

3.1 Point estimation

We consider local polynomial estimation which is common in the nonparametric regression literature. In particular, we focus on the two most popular versions, local constant and local linear estimation. The local constant estimator is also called Nadaraya-Watson estimator (Nadaraya, 1964; Watson, 1964). For this model, it is found by minimizing the following weighted sum of squares with respect to $m(\cdot)$:

$$SSR_{lc}(\tau) = \sum_{t=1}^{n} \{y_t - m(t/n)\}^2 K\left(\frac{t/n - \tau}{h}\right), \qquad \text{for } \tau \in (0,1), \tag{3.1}$$

where $K(\cdot)$ is a kernel function and $h > 0$ is a bandwidth, which should satisfy Assumptions 4 and 5 given below.

The solution to the above minimization problem can be expressed as

$$\hat{m}(\tau) = (nh)^{-1} \sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right) y_t, \qquad \text{for } \tau \in (0,1)\,. \tag{3.2}$$

While the Nadaraya-Watson estimator locally approximates the trend function by a constant function, the second estimation method we consider locally fits a linear function to the data around a given point to obtain an estimate of the trend function at this point. We need to minimize the following weighted sum of squares with respect to $m(\cdot)$ and $m^{(1)}(\cdot)$ to obtain this estimator for a given point $\tau \in (0,1)$

$$SSR_{ll}(\tau) = \sum_{t=1}^{n} \left\{ y_t - m(t/n) - m^{(1)}(t/n)\,(t/n - \tau)) \right\}^2 K\left(\frac{t/n - \tau}{h}\right), \qquad \text{for } \tau \in (0,1)\,. \tag{3.3}$$

To obtain the solution to this problem, define

$$\mathbf{x}_t(\tau) \equiv \begin{pmatrix} 1 \\ t/n - \tau \end{pmatrix}.$$

Then,

$$\begin{pmatrix} \hat{m}(\tau) \\ \hat{m}^{(1)}(\tau) \end{pmatrix} = \left(\sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right) \mathbf{x}_t(\tau)\mathbf{x}_t(\tau)' \right)^{-1} \sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right) \mathbf{x}_t(\tau) y_t \tag{3.4}$$

solves the minimization problem corresponding to local linear estimation. The local linear estimator is shown to be superior to the local constant estimator at points which are close to the boundaries of the sample. At these points, the local constant estimator suffers from boundary effects which the local linear estimator does not (see e.g. Cai (2007) and Fan (1992) for an extensive comparison).

Assumption 4. *$K(\cdot)$ is a probability density, symmetric, twice continuously differentiable and with compact support. We further assume that $\int_{\mathbb{R}} K(\eta)^4 d\eta < \infty$.*

Assumption 5. *For the bandwidth, we require $h = h(n) = o(1)$ as well as $nh^2 \to \infty$ as $n \to \infty$.*

Assumptions 4 and 5 are standard assumptions in the nonparametric kernel smoother literature. Most frequently used kernels satisfy Assumption 4.

Remark 1. *Our estimators can easily be generalized to account for missing data. Assume we observe data on times t_i for $i = 1, \ldots, n$, where $t_1, \ldots, t_n$ are not necessarily equally spaced. The local constant estimator then is*

$$\hat{m}(\tau) = (nh)^{-1} \sum_{i=1}^{n} K\left(\frac{t_i/t_n - \tau}{h}\right) y_{t_i},$$

while the local linear estimator can be adjusted to

$$\begin{pmatrix} \hat{m}(\tau) \\ \hat{m}'(\tau) \end{pmatrix} = \left(\sum_{i=1}^{n} K\left(\frac{t_i/t_n - \tau}{h}\right) \mathbf{x}_{t_i}(\tau)\mathbf{x}_{t_i}(\tau)' \right)^{-1} \sum_{i=1}^{n} K\left(\frac{t_i/t_n - \tau}{h}\right) \mathbf{x}_{t_i}(\tau) y_{t_i},$$

where $\mathbf{x}_{t_i}(\tau) \equiv (1, t_i/t_n - \tau)'$. *To lighten the notational burden, we use the equally spaced time index* $t = 1, ..., n$ *in the remainder of the paper, although we will indicate how methods and analysis can be extended to missing data in further remarks when appropriate.*

The bandwidth parameter h plays an important role for both estimators. It is also called smoothing parameter; large bandwidths produce a very smooth estimate, while for small bandwidth, the trend estimate becomes rougher. Bandwidth selection is a crucial aspect of the estimation procedure which is usually done based on the data. Leave-one-out cross-validation is the most popular data-based method for bandwidth selection, but shows certain problems when applied to time series data, as it is designed for independent observations. Chu and Marron (1991) show that in the presence of positive correlation, this criterion systematically selects very small bandwidths, producing estimates which are too rough. With negative correlation, the chosen bandwidth will be large and the estimate too smooth. Therefore, Chu and Marron (1991) propose to use a time series version of such a criterion, called modified cross-validation (MCV). It is based on the following criterion:

$$CV_k(h) = \frac{1}{n} \sum_{t=1}^{n} (\hat{m}_{k,h}(t/n) - \hat{y}_t) \tag{3.5}$$

where

$$\hat{m}_{k,h}(\tau) = \frac{(n-2k-1)^{-1} \sum_{t:|t-\tau n|>k} K\left(\frac{t/n-\tau}{h}\right) \hat{y}_t}{(n-2k-1)^{-1} \sum_{t:|t-\tau n|>k} K\left(\frac{t/n-\tau}{h}\right)} \tag{3.6}$$

is a leave-($2k$+1)-out version of the leave-one-out estimator of ordinary cross-validation, which leaves out the observation receiving the highest weight. We recommend using this selection criterion or other modifications of cross-validation which are applicable in a time series setting. In addition, as a kind of robustness check we suggest to run the estimation using a range of different bandwidths and to visually inspect the resulting trend estimates to check if the most prominent patterns are visible with all estimates.

3.2 Bootstrap confidence intervals

To generate confidence intervals around the trend estimate, we propose a modified version of the wild bootstrap, which was originally designed to handle heteroskedastic data. We refer the reader to Davidson and Flachaire (2008) for a detailed overview of the wild bootstrap in a linear regression framework. The general idea of this method is to generate bootstrap errors as

$$z_t^* = \xi_t^* \hat{z}_t,$$

where $\hat{z}_t$ denote the residuals of the nonparametric trend model. In the ordinary wild bootstrap, the random variables $\{\xi_t^*\}_{t=1}^n$ are i.i.d. and thus, the dependence structure present in the residuals does not get correctly reflected in the bootstrap errors. To overcome this drawback and to mimic the dependence, $\{\xi_t^*\}_{t=1}^n$ is allowed to be dependent in the autoregressive wild bootstrap. Specifically, we generate $\nu_1^*, ..., \nu_n^*$ as i.i.d. $\mathcal{N}(0, 1-\gamma^2)$ and let

$$\xi_t^* = \gamma \xi_{t-1}^* + \nu_t^* \qquad t = 2, ..., n \qquad \xi_1^* \sim \mathcal{N}(0,1)$$

with $\gamma = \gamma(n)$. More specifically, in line with Smeekes and Urbain (2014) we let $\gamma = \theta^{1/\ell}$ such that ℓ satisfies Assumption 6 and θ is a fixed parameter. With this specification ℓ can be interpreted in a similar way as the block length parameter in a block bootstrap and its choice constitutes a trade-off between capturing more of the dependence present in the residuals with a large value of the tuning parameter and allowing for more variation in the bootstrap samples with a smaller value for l. Larger variation could lead to a better approximation of the sampling distribution.

Assumption 6. *For the tuning parameter ℓ of the autoregressive wild bootstrap $l \to \infty$ as $n \to \infty$ should hold. Further we assume that $\ell = o\left(\min\left\{(nh)^{1/2}, \tilde{h}^{-2}, (n\tilde{h})^{1/2}\right\}\right)$.*

The bootstrap algorithm consists of three steps and is described as follows:

Step 1 Estimate model (2.1) and form a residual series. This means, calculate

$$\hat{z}_t = y_t - \tilde{m}(t/n), \qquad t = 1, ..., n,$$

where the estimate $\tilde{m}(\tau)$ is obtained by using bandwidth $\tilde{h}$, which does not have to be equal to h.

Step 2 Generate $\nu_1^*, ..., \nu_n^*$ as i.i.d. $\mathcal{N}(0, 1-\gamma^2)$ and let

$$\xi_t^* = \gamma \xi_{t-1}^* + \nu_t^* \qquad t = 2, ..., n \qquad \xi_1^* \sim \mathcal{N}(0,1)$$

with $\gamma = \gamma(n)$.

Step 3 Calculate the bootstrap errors z_t^* as

$$z_t^* = \xi_t^* \hat{z}_t,$$

Now, generate the bootstrap observations by

$$y_t^* = \tilde{m}(t/n) + z_t^*, \qquad t = 1, ..., n,$$

where $\tilde{m}(t/n)$ is the same estimate as in the first step.

Steps 2 and 3 have to be repeated B times to construct the B bootstrap series $\{y_t^*\}_{t=1}^n$. Note that in Step 1 of the above algorithm, it can be useful perform the nonparametric estimation with a different bandwidth compared to the original estimation. In general, this bandwidth is assumed to satisfy the following assumption:

Assumption 7. *Similarly to Assumption 5, for the bandwidth $\tilde{h}$, we have that $\tilde{h} = \tilde{h}(n) = o(1)$ as well as $n\tilde{h}^5 \to \infty$ as $n \to \infty$.*

We follow the recommendation of Bühlmann (1998) and suggest to use $\tilde{h} = Ch^{5/9}$, with $C > 0$. Compared to the original bandwidth h, the bandwidth $\tilde{h}$ is larger. This produces an oversmoothed estimate as starting point for the bootstrap procedure. The reason why we recommend oversmoothing is the presence of the asymptotic bias. As seen in Theorem 4.1, the limiting normal distribution of the nonparametric estimator contains a bias term which includes the second derivative of the trend function. This can only be consistently estimated using a larger bandwidth $\tilde{h}$ such that $\tilde{h}n^{1/5} \to \infty$. This is derived in Gasser and Müller (1984) and explained in more detail in Bühlmann (1998). The limiting distribution of the bootstrap estimator should as accurately as possible approximate the true limiting distribution, therefore the asymptotic bias has to be consistently estimated in the procedure. Formally, this means that the following statement holds

$$(nh)^{1/2} \left(\mathbb{E}^* \left[\hat{m}^*(\tau)\right] - \tilde{m}(\tau)\right) - B_{as}(\tau) = o_p(1),$$

for $\tau \in (0, 1)$, which says that the bias in the bootstrap converges to the true asymptotic bias.

Remark 2. *The bootstrap can handle missing data without further modifications whenever we have an underlying regular frequency on which we could observe data that we can simulate the bootstrap random variables on. For example, in our empirical application in Section 6, the underlying frequency of atmospheric ethane measurements is daily. Assume we observe data on times $t_1, \dots, t_n$ that form a subset of the underlying observational times $s = 1, \dots, N$. We simulate $\{\xi_s^*\}_{s=1}^N$. Subsequently we only use the subset that corresponds to the actually observed data points $\left\{\xi_{t_i}^*\right\}_{i=1}^n$ and obtain $z_{t_i}^* = \xi_{t_i}^* \hat{z}_{t_i}$ for $i = 1, \dots, n$. This way the missing data structure is preserved in the bootstrap sample, while the correlation between observations t_j and t_{j+1} is governed only by their distance $t_{j+1} - t_j$, irrespective of whether the observations in between are observed or not, which ensures a coherent bootstrap sample.*

Pointwise confidence intervals, $I_{\alpha,p}$, for a parameter curve $m(\cdot)$ are of the form

$$\lim_{n\to\infty} \left[\mathbb{P}\left(m(t/n) \in I_{\alpha,p}(t/n)\right)\right] \geq 1 - \alpha \qquad \text{for } t = 1, ..., n. \tag{3.7}$$

To construct these intervals for $m(\cdot)$, the centered quantity $\hat{m}^*(t/n) - \tilde{m}(t/n)$ is needed. From there it is straightforward to determine pointwise two-sided confidence intervals for a confidence level of $(1 - \alpha)$. These are exactly the values for every t, between which $(1 - \alpha)$ of the bootstrap deviations fall. Formally, this can be stated as

$$I_{\alpha,p}(t/n) = \left[\hat{m}(t/n) - \hat{q}_{1-\alpha/2}, \hat{m}(t/n) - \hat{q}_{\alpha/2}\right], \tag{3.8}$$

where $(1 - \alpha)$ is the confidence level and $\hat{q}_\alpha = \inf\left\{u; \mathbb{P}^* \left[\hat{m}^*(t/n) - \tilde{m}(t/n) \leq u\right] \geq \alpha\right\}$. The subscript p stands for pointwise to distinguish $I_{\alpha,p}$ from the simultaneous counterpart I_α. The confidence intervals which are determined by the $(\alpha/2)$-quantile and $(1 - \alpha/2)$-quantile of the bootstrap distribution yield asymptotically correct $(1-\alpha)$ pointwise confidence intervals according to Theorem 4.2. A link over time cannot be established with these intervals. As argued above, simultaneous confidence bands are more informative. We use the remainder of this section to formalize this idea and to describe how these bands can be obtained from pointwise confidence intervals.

Using the same notation as above, the following bands are simultaneous over the set G

$$\lim_{n\to\infty}\left[\mathbb{P}\left(m(t/n)\in I_\alpha(t/n)\ \text{ for }\ t\in G\right)\right]\geq 1-\alpha. \tag{3.9}$$

Practical implementation of the simultaneity follows a three-step procedure which was first presented in this context in Bühlmann (1998). It is a search algorithm based on the ordered deviations, $\hat{m}^*(t/n)-\tilde{m}(t/n)$, of bootstrap estimates from the original estimate. The first step is to construct pointwise quantiles from the deviations in the same way as above:

Step 1 Obtain pointwise quantiles for varying α_p - starting by $1/B$ and ending in α:

$$\hat{q}_{\alpha_p/2}(t/n),\hat{q}_{1-\alpha_p/2}(t/n),\quad t\in G,$$

where $\hat{q}_{\alpha_p}=\inf\left\{u;\mathbb{P}^*\left[\hat{m}^*(t/n)-\tilde{m}(t/n)\leq u\right]\geq\alpha_p\right\}$ is a pointwise quantile.

Step 2 Choose α_s as

$$\alpha_s=\text{argmin}_{1/B\leq\alpha_p\leq\alpha}\left|\mathbb{P}^*\left[\hat{q}_{\alpha_p/2}(t/n)\leq\hat{m}^*(t/n)-\tilde{m}(t/n)\leq\hat{q}_{1-\alpha_p/2}(t/n);t\in G\right]-(1-\alpha)\right|$$

Step 3 Construct the simultaneous confidence bands as

$$I_{\alpha_s}(t/n)=\left[\hat{m}(t/n)-\hat{q}_{1-\alpha_s/2}(t/n),\hat{m}(t/n)-\hat{q}_{\alpha_s/2}(t/n)\right]\quad t\in G.$$

In the second step of this procedure, a pointwise error α_s is found for which a fraction of approximately $(1-\alpha)$ of all centered bootstrap estimates falls within the resulting confidence intervals, for all points of the set G. An important aspect to stress at this point is that the coverage needs to be seen over the set G and not point by point. As soon as an estimated bootstrap deviation falls outside the given intervals at one point within the set, it is not counted for the probability in Step 2. The confidence intervals with pointwise coverage $(1-\alpha_s)$ become simultaneous confidence bands with coverage $(1-\alpha)$.

Remark 3. *While the confidence bands constructed in this way are of variable size, Neumann and Polzehl (1998) consider simultaneous confidence bands of uniform size as an alternative approach. They are of the form*

$$I^*_\alpha(t/n)=\left[\hat{m}(t/n)-t^*_{1-\alpha},\hat{m}(t/n)+t^*_{1-\alpha}\right].$$

The quantile $t^*_{1-\alpha}$ is determined as the $(1-\alpha)$-quantile of the distribution of the quantity

$$U^*_n=\sup_t\left\{\left|\hat{m}^*(t/n)-\tilde{m}(t/n)\right|\right\}.$$

Although for these intervals simultaneity can be established over the whole sample, we do not go further down this path because we prefer to construct confidence bands with non-equal width. They offer the advantage that at points with more variability the confidence bands are wider and become more narrow for periods with less variability, which we consider to be a valuable feature which the alternative approach cannot achieve.

4 Asymptotic Theory

We first provide the pointwise limiting normal distribution of the local constant estimator $\hat{m}(\cdot)$. While a similar result is derived in Bühlmann (1998), we present this result again as the presence of non stationary volatility requires a small modification.

Theorem 4.1. *Under Assumptions 1-5, for any $\tau \in (0,1)$, we have as $n \to \infty$:*

$$\sqrt{nh}\left(\hat{m}(\tau) - m(\tau)\right) \xrightarrow{d} \mathcal{N}\left(B_{as}(\tau), \sigma^2_{as}(\tau)\right),$$

where

$$\begin{aligned} B_{as}(\tau) &= \lim_{n\to\infty} \sqrt{nh}h^2/2m^{(2)}(\tau)\int_{\mathbb{R}} u^2K(u)du \\ \sigma^2_{as}(\tau) &= \sigma(\tau)^2\left[R(0) + 2\sum_{m=1}^{\infty} R(m)\right]\int_{\mathbb{R}} K^2(u)du. \end{aligned}$$

The asymptotic variance, $\sigma^2_{as}(\tau)$, has a similar form as in Bühlmann (1998) and coincides with $\sigma^2_{as}(\tau) = 2\pi\sigma(\tau)^2 f_u(0)\int_{\mathbb{R}} K^2(u)du$, where $f_u(\lambda) = (2\pi)^{-1}\sum_{k=-\infty}^{\infty} Cov(u_0, u_k)e^{-i\lambda k}$ is the spectral density of the noise process $\{u_t\}_{t=1}^n$. The only difference to the setting of Bühlmann (1998) is the presence of $\sigma(\tau)$ which appears due to the fact that we allow for a non-constant variance.

The second theorem is the bootstrap analogue of Theorem 4.1 and it establishes consistency of the autoregressive wild bootstrap method in this new setting for the local constant estimator.

Theorem 4.2. *Under Assumptions 1-7, for any $\tau \in (0,1)$, we have as $n \to \infty$:*

$$\sqrt{nh}\left(\hat{m}^*(\tau) - \tilde{m}(\tau)\right) \xrightarrow{d^*}_p \mathcal{N}\left(B_{as}(\tau), \sigma^2_{as}(\tau)\right),$$

with $B_{as}(\tau)$ and $\sigma^2_{as}(\tau)$ as given in Theorem 4.1. Therefore, it holds that, as $n \to \infty$

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}^*\left[\sqrt{nh}\left(\hat{m}^*(\tau) - \tilde{m}(\tau)\right) \leq x\right] - \mathbb{P}\left[\sqrt{nh}(\hat{m}(\tau) - m(\tau)) \leq x\right]\right| = o_p(1).$$

The pointwise validity of the bootstrap confidence intervals in the sense of (3.7) follows directly from this pointwise convergence result. The bias term $B_{as}(\tau)$ of the asymptotic distribution contains the second derivative of the trend function, which is the reason why we need oversmoothing in the bootstrap as argued in the previous section. We now move to uniform behavior of the bootstrap.

A crucial step for the validity of simultaneous confidence bands as in (3.9) is to consider what happens within $h(n)$-neighborhoods around time points τ. The estimates $\hat{m}(\tau)$ and $\hat{m}(\vartheta)$ are asymptotically independent for $\tau \neq \vartheta$ being two distinct time points. This is not the case when the distance between τ and ϑ is of order $h(n)$. For such points, the estimators show a non-zero correlation. To be able to show validity of the neighborhood simultaneity of the confidence bands constructed using the wild bootstrap scheme, we have to establish that the correlation is correctly reflected in the bootstrap distribution. This is done in the next theorem. Before we state the theorem, we need to define the following quantities, for a time point $0 < \tau_0 < 1$:

$$\begin{aligned} N_{\tau_0,n}(\tau) &= \sqrt{nh}\left(\hat{m}(\tau_0+\tau h) - m(\tau_0+\tau h)\right) \\ N^*_{\tau_0,n}(\tau) &= \sqrt{nh}\left(\hat{m}^*(\tau_0+\tau h) - \tilde{m}(\tau_0+\tau h)\right). \end{aligned}$$

For $-1 \leq \tau \leq 1$, the quantities $N_{\tau_0,n}(\tau)$ and $N^*_{\tau_0,n}(\tau)$ define $h(n)$-neighborhoods around τ_0 and are the objects of interest in Theorem 4.3. Since $h(n) = o(1)$, we assume without loss of generality that $m(\tau_0+\tau h)$ is always defined. We denote the space of continuous real-valued functions on $[-1,1]$ by $\mathcal{C}\,[-1,1]$. In the following theorem $\Rightarrow$ stands for weak convergence in $\mathcal{C}\,[-1,1]$ with respect to the sup-norm.

Theorem 4.3. *Under Assumptions 1-7, we have for any* $0 < \tau_0 < 1$

$$\begin{aligned} &\left\{N_{\tau_0,n}(\tau) - B_{as}(\tau_0)\right\}_{\tau\in[-1,1]} \Rightarrow \left\{W(\tau)\right\}_{\tau\in[-1,1]}, \\ &\left\{N^*_{\tau_0,n}(\tau) - B_{as}(\tau_0)\right\}_{\tau\in[-1,1]} \Rightarrow \left\{W(\tau)\right\}_{\tau\in[-1,1]} \quad \textit{in probability,} \end{aligned}$$

where $\{W(\tau)\}_{\tau\in[-1,1]}$ *is a Gaussian process with*

$$\begin{aligned} \mathbb{E}(W(\tau)) &= 0 \\ Cov(W(\tau), W(\vartheta)) &= \sigma(\tau)^2\left[R(0) + 2\sum_{m=1}^{\infty} R(m)\right]\int_{\mathbb{R}} K(\omega)K(\omega+\vartheta-\tau)d\omega. \end{aligned}$$

This theorem describes the uniform behavior of the bootstrap within an $h(n)$-neighborhood around a point τ_0 and establishes that it approximates the true uniform behavior of the asymptotic distribution correctly. More specifically, the set G introduced in the previous section can contain a finite number of neighborhoods $U_1, ..., U_k$ of lengths $ch(n)$, $0 < c < \infty$, and a union thereof. Validity of the bootstrap to create confidence bands that are simultaneous over such a set will be established in the next section. Simultaneity over such a set might be less informative than simultaneity over the whole sample, it can nevertheless be of great interest in applications. For example, constructing confidence bands with simultaneous coverage over two time periods - one located early in the sample and the other one at the end - is useful when judging if there was an upward (or downward) movement of the trend at the end of the time period when compared to the beginning. In this way we can draw conclusions about the development over the covered time stretch, which would not be possible with pointwise confidence intervals. Theorem 4.3 directly implies the following corollary concerning a finite union of such neighborhoods:

Corollary 4.4. *Under Assumptions 1-7, let* $0 < \tau_1 < \cdots < \tau_k < 1$ *and* $U_i = [\tau_i - h, \tau_i + h]$*, for* $i = 1, ..., k < \infty$*. Then, for* $\mathbf{x} \in \mathbb{R}$ *and* $n \to \infty$*,*

$$\mathbb{P}^* \left[\max_{1 \le i \le k} \sup_{\tau \in U_i} \sqrt{nh}(\hat{m}^*(\tau) - \tilde{m}(\tau)) \le \mathbf{x} \right] - \mathbb{P} \left[\max_{1 \le i \le k} \sup_{\tau \in U_i} \sqrt{nh}(\hat{m}(\tau) - m(\tau)) \le \mathbf{x} \right] = o_p(1).$$

This result naturally follows from Theorem 4.3. The proof is a straight-forward extension of the proof of this theorem.

Remark 4. *While we for notational simplicity only consider equally spaced data explicitly in the asymptotic analysis, we can extend our analysis to missing data. As in Remark 2, assume we observe data on a subset* $t_1, \dots, t_n$ *of* $1, \dots, N$*. Assume that* $n/N/rightarrowc > 0$*. If the missing data are not pervasive asymptotically, that is when* $c = 1$*, our asymptotic analysis can be extended by simply adjusting the indices used for estimators and bootstrap as in Remarks 1 and 2. For the more interesting case where* $c < 1$ *such that missing values remain present asymptotically, we conjecture that our analysis can be adapted to show bootstrap validity, although depending on the assumed missing data generating mechanism, our results may have to be restricted to* $\tau \in (0,1)\mathcal{M}$*, where* $\mathcal{M}$ *denotes the set of* τ *for which no (or too few) observations are observed. However, as this case presents significant notational complexities, without adding much to the understanding of the method and its properties, we do not consider this case explicitly in the paper.*

Remark 5. *As discussed in Bühlmann (1998, p. 53), our analysis for the local constant estimator extends directly to higher order local polynomial estimators such as the local linear estimator. However, notation becomes significantly more cumbersome, so we focus on the local constant estimator for the asymptotic analysis.*

5 Simulation study

For the simulation exercise, we simulate time series with a trending behavior which could appear in climatological time series. Consider the following data generating process, which is a smooth transition version of a broken trend model with one break occurring at time c.

$$y_t = \beta t \left(1 - G(t, \lambda, c)\right) + \delta_{t,c}(t - c)G(t, \lambda, c) + \sigma_t u_t \tag{5.1}$$

where

$$\delta_{t,c} = \begin{cases} 0 & \text{if } t \le c, \\ \delta & \text{if } t > c \end{cases}$$

and for $\lambda > 0$,

$$G(t, \lambda, c) = \left(1 + \exp\left\{-\lambda(t/n - c)\right\}\right)^{-1}.$$

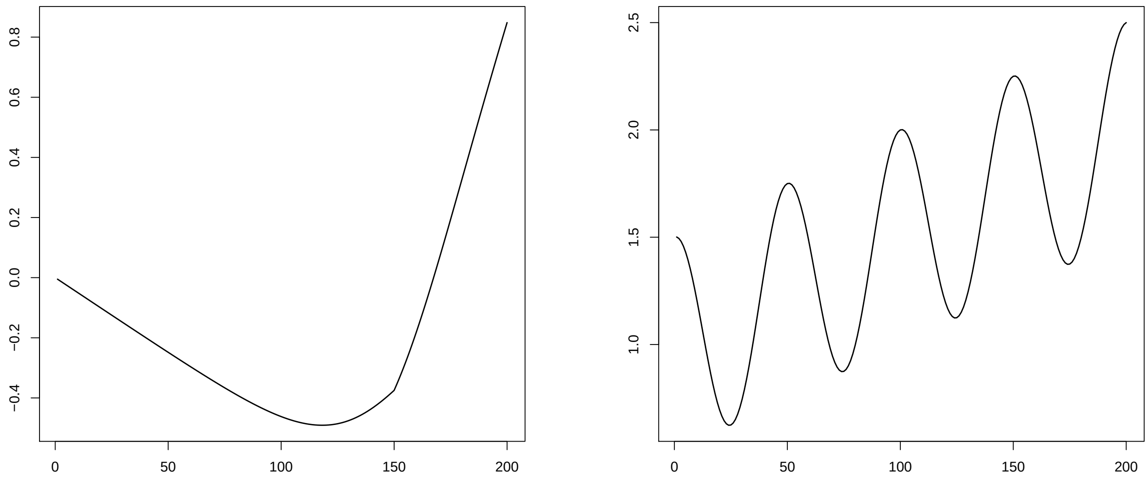
The error term $\{u_t\}_{t=1}^n$ follows an ARMA(1,1) model

$$u_t = \phi u_{t-1} + \psi \epsilon_{t-1} + \epsilon_t \qquad\qquad \epsilon_t \sim \mathcal{N}\left(0, \frac{(1-\phi^2)/4}{1+\psi^2-2\phi\psi}\right),$$

where we vary the parameters ϕ and ψ to investigate the impact of serial correlation on our method. The variance of ϵ_t is normalized such that the signal to noise ratio does not depend on the specific choice of the AR and MA parameter. Furthermore, we introduce heteroskedasticity with the process $\{\sigma_t\}_{t=1}^n$. We consider two scenarios, first we choose σ_t to be constant over time and second, we use the volatility process

$$\sigma_t = \sigma_0 + (\sigma_* - \sigma_0)(t/n) + a\cos(2\pi k(t/n)). \tag{5.2}$$

Model (5.1) is a shifting mean model as considered by Gonzalés and Teräsvirta (2008). The function $G((t,\lambda,c)$ is the transition function with time as transition variable. Its inputs apart from time are the location of the shift - the parameter c - as well as the smoothness of the shift, determined by λ. For large values of λ the shift happens almost instantaneous, while it is more smooth for smaller values of this parameter. In our simulations, we fix $\lambda = 10$. The other parameters of our specific data generating process will be chosen in such a way that the time series experiences a downward trend during the first three quarters which turns into a steeper upward trend in the last quarter.



(a) Simulated trend function

(b) Simulated variance process

Figure 1: Visualizing the DGP

This mimics the general pattern which is expected to occur in atmospheric ethane time series and therefore fits our application well. More specifically, this means we set the location of the shift to occur at $c = 0.75$. The slope of the trend gradually changes from $\beta = -1$ before the shift to $\beta + \delta = 3$ after the shift. This is illustrated in Figure 1(a). For the variance process, inspired by the empirical application, we consider a cyclical component with trend. We have to choose four parameters: the start and end point of the trend - σ_0 and σ_* - as well as the specifics of the cyclical component. The parameter a fixes the amplitude of the cycle while k determines how many cycles there are. We set $\sigma_0 = 1$, $\sigma_* = 2$, $a = 0.5$ and $k = 4$. This process is displayed in Figure 1(b).

While we fix the parameters that enter the transfer function and the variance process, we consider different degrees of dependence by varying the AR and MA parameters. For every parameter specification, we run 1000 Monte Carlo simulations. In the estimation step, we apply the local constant estimator based on the Epanechnikov kernel which is given by the function

$$K(x) = \frac{3}{4}(1 - x^2)\mathbb{1}_{\{|x| \leq 1\}}.$$

We consider a range of different bandwidth parameters. Specifically, we use $h = 0.02, 0.04, 0.06, 0.08$. In the first step of the bootstrap procedure oversmoothing is used with $\tilde{h} = Ch^{5/9}$ and $C = 2$. In the second step, we consider different values for the AR parameter γ, where the value $\gamma = 0$ reduces the autoregressive wild bootstrap to a wild bootstrap without autoregressive element. For each specification, we run 1000 Monte Carlo simulations.

We report average pointwise as well as simultaneous coverage for a sample size of $n = 200$, based on $B = 599$ bootstrap replications. The nominal coverage is 95%. We also report the average median length of the confidence intervals in parenthesis underneath the respective coverage. For simultaneous coverage, the trend curve has to lie within the confidence bands for all points of the set G. There are two sets of points we consider, G and G_{sub}. They consist of four and two intervals, respectively, whose length is approximately equal to $2h$. As in Bühlmann (1998); we take

$$G = \bigcup_{i=1}^{4} U_i(h),$$
$$G_{sub} = U_1(h) \cup U_4(h),$$

with

$$U_i(h) = \{x_i - h + j/100;\;\; j = 0, ..., [200h]\}\,, \qquad x_i = i/5.$$

We first report results for the homoskedastic case. The results on pointwise coverage are given in Table 1, while Tables 2 and 3 show simultaneous coverage probabilities for the two sets G and $G.sub$, respectively. The tables consist of four main blocks, one for each bandwidth. Within the blocks, the individual rows contain results for different combinations of AR and MA parameters. We consider seven such different combinations, independence, two AR models with positive correlation, two AR models with negative correlation and two MA models. For each of them, we vary γ from 0 to 0.6 in steps of 0.2.

Table 1 shows that the bootstrap confidence intervals provide good pointwise coverage. For the independent case and the cases with negative or small positive correlation, the coverage probabilities are close to the nominal level. The same holds for the case with strong negative correlation. The only specifications for which the coverage is clearly below the nominal level are the ones with $\phi = 0.5$ and $\psi = 0.5$. In these cases, the data deviates from the trend line in clusters due to the strong positive correlation. This causes the nonparametric estimate to go through these clusters and thus, to deviate significantly from the true trend. The confidence bands are in these situations not wide enough to cover the true trend. More conservative confidence intervals would be needed.

Concerning the autoregressive parameter of the wild bootstrap, we can observe that whenever the data is serially correlated, the autoregressive wild bootstrap ($\gamma \neq 0$) provides better coverage than the standard wild bootstrap ($\gamma = 0$). In addition, with stronger correlation, a larger value for γ should be preferred. This is partly supported by our results. The case $\gamma = 0.6$, however, provides consistently lower coverage. Given that γ should increase with the sample size, this value seems to be too large for the small sample size. In general, the coverage probabilities do not vary substantially with the autoregressive parameter and therefore, the sensitivity to this parameter is limited.

When we look at the different blocks, the bootstrap shows similar coverage independent of the value we select for the bandwidth parameter. Since bandwidth selection plays an important role in nonparametric estimation and there is no optimal bandwidth selection method in most applications, this is an important finding.

We can observe a similar pattern in Tables 2 and 3; while the overall coverage is lower for the set G than for $G.sub$, the results seem to be robust with respect to the bandwidth. This is not surprising, since the former set covers twice as many points as the latter. Additionally, the confidence bands are consistently more narrow with G than they are with $G.sub$. This could be due to the fact that the median is taken over a larger set with G than with $G.sub$. Similar to the pointwise coverage results, the simultaneous coverage is close to the nominal level for the two cases with negative correlation as well as the independent case. Weak positive correlation can also be handled decently. Problems arise when $\phi = 0.5$ or $\psi = 0.5$. Then the coverage drops to around 60% for G and 70% for $G.sub$.

Second, we show results for the heteroskedastic case in Tables 4 and 5. Given the previous findings, we restrict ourselves to one bandwidth and consider a limited number of AR and MA models. We also drop the case $\gamma = 0.6$. Both sets of results indicate that the autoregressive wild bootstrap can handle heteroskedasticity well. The confidence intervals become wider to account for the added difficulty, but the coverage probabilities do not differ much from the homoskedastic case. This shows that, indeed, our method seems to be robust to heteroskedasticity as the coverage is not affected by it.

6 Application to atmospheric ethane

We use our methodology to investigate the trending behavior of a time series of atmospheric ethane emissions which is derived from observations performed at the Jungfraujoch station in the Swiss Alps. This station can be found on the saddle between the Jungfrau and the Mönch, located at 46.55° N, 7.98° E, 3580 m altitude. Ethane is the most abundant hydrocarbon gas in the atmosphere after methane and it is used as a measure of atmospheric pollution. It contributes to the formation of ground-level ozone and it influences the lifetime of methane which classifies it as an indirect greenhouse gas. This series has been studied in Franco et al. (2015) and it is available from the Network for the Detection of Atmospheric Composition Change website at

ARMA(ϕ,ψ)	$\gamma=0$	$\gamma=0.2$	$\gamma=0.4$	$\gamma=0.6$	$\gamma=0$	$\gamma=0.2$	$\gamma=0.4$	$\gamma=0.6$
	$h=0.02$				$h=0.04$			
$\phi=0,\ \psi=0$	0.947	0.952	0.947	0.914	0.948	0.950	0.944	0.914
	(0.518)	(0.505)	(0.467)	(0.398)	(0.397)	(0.393)	(0.373)	(0.331)
$\phi=0.2,\ \psi=0$	0.904	0.920	0.917	0.879	0.902	0.911	0.910	0.881
	(0.501)	(0.501)	(0.476)	(0.420)	(0.386)	(0.395)	(0.385)	(0.354)
$\phi=0.5,\ \psi=0$	0.793	0.828	0.840	0.806	0.782	0.816	0.830	0.810
	(0.425)	(0.447)	(0.442)	(0.409)	(0.336)	(0.363)	(0.371)	(0.357)
$\phi=-0.2,\ \psi=0$	0.971	0.972	0.966	0.937	0.973	0.972	0.966	0.939
	(0.507)	(0.485)	(0.436)	(0.363)	(0.391)	(0.377)	(0.346)	(0.300)
$\phi=-0.5,\ \psi=0$	0.989	0.989	0.984	0.964	0.989	0.989	0.986	0.969
	(0.443)	(0.412)	(0.360)	(0.287)	(0.349)	(0.327)	(0.288)	(0.239)
$\phi=0,\ \psi=0.2$	0.914	0.926	0.924	0.890	0.913	0.921	0.919	0.890
	(0.503)	(0.503)	(0.476)	(0.418)	(0.388)	(0.395)	(0.385)	(0.351)
$\phi=0,\ \psi=0.5$	0.870	0.895	0.899	0.867	0.871	0.892	0.897	0.873
	(0.453)	(0.467)	(0.455)	(0.409)	(0.354)	(0.374)	(0.374)	(0.349)
	$h=0.06$				$h=0.08$			
$\phi=0,\ \psi=0$	0.943	0.944	0.940	0.917	0.933	0.936	0.932	0.916
	(0.340)	(0.340)	(0.330)	(0.304)	(0.306)	(0.311)	(0.308)	(0.294)
$\phi=0.2,\ \psi=0$	0.899	0.910	0.910	0.886	0.892	0.902	0.904	0.888
	(0.331)	(0.343)	(0.341)	(0.324)	(0.299)	(0.313)	(0.317)	(0.309)
$\phi=0.5,\ \psi=0$	0.786	0.821	0.835	0.825	0.784	0.822	0.836	0.833
	(0.293)	(0.321)	(0.334)	(0.330)	(0.267)	(0.296)	(0.313)	(0.315)
$\phi=-0.2,\ \psi=0$	0.966	0.966	0.961	0.940	0.955	0.955	0.951	0.000
	(0.335)	(0.327)	(0.306)	(0.278)	(0.302)	(0.299)	(0.289)	(0.000)
$\phi=-0.5,\ \psi=0$	0.980	0.981	0.979	0.965	0.967	0.968	0.967	0.956
	(0.304)	(0.287)	(0.259)	(0.231)	(0.277)	(0.267)	(0.253)	(0.246)
$\phi=0,\ \psi=0.2$	0.910	0.918	0.917	0.896	0.902	0.911	0.911	0.898
	(0.332)	(0.343)	(0.342)	(0.320)	(0.300)	(0.314)	(0.318)	(0.307)
$\phi=0,\ \psi=0.5$	0.873	0.893	0.898	0.883	0.869	0.887	0.894	0.886
	(0.306)	(0.327)	(0.334)	(0.319)	(0.278)	(0.301)	(0.312)	(0.306)

Table 1: Pointwise coverage probabilities

ARMA(ϕ,ψ)	$\gamma = 0$	$\gamma = 0.2$	$\gamma = 0.4$	$\gamma = 0.6$	$\gamma = 0$	$\gamma = 0.2$	$\gamma = 0.4$	$\gamma = 0.6$
	$h = 0.02$				$h = 0.04$			
$\phi = 0$, $\psi = 0$	0.891	0.887	0.868	0.779	0.884	0.896	0.856	0.737
	(0.378)	(0.543)	(0.341)	(0.291)	(0.285)	(0.283)	(0.268)	(0.238)
$\phi = 0.2$, $\psi = 0$	0.826	0.828	0.795	0.688	0.779	0.775	0.781	0.646
	(0.366)	(0.366)	(0.348)	(0.307)	(0.278)	(0.284)	(0.277)	(0.255)
$\phi = 0.5$, $\psi = 0$	0.518	0.617	0.630	0.503	0.406	0.518	0.527	0.477
	(0.310)	(0.327)	(0.324)	(0.300)	(0.241)	(0.261)	(0.267)	(0.258)
$\phi = -0.2$, $\psi = 0$	0.939	0.918	0.909	0.815	0.962	0.944	0.927	0.847
	(0.371)	(0.354)	(0.319)	(0.265)	(0.281)	(0.271)	(0.249)	(0.216)
$\phi = -0.5$, $\psi = 0$	0.986	0.978	0.957	0.886	0.994	0.992	0.986	0.948
	(0.323)	(0.301)	(0.262)	(0.210)	(0.251)	(0.235)	(0.206)	(0.172)
$\phi = 0$, $\psi = 0.2$	0.820	0.841	0.806	0.697	0.808	0.800	0.781	0.674
	(0.367)	(0.368)	(0.348)	(0.306)	(0.278)	(0.284)	(0.277)	(0.253)
$\phi = 0$, $\psi = 0.5$	0.696	0.765	0.745	0.628	0.671	0.728	0.698	0.622
	(0.331)	(0.342)	(0.333)	(0.300)	(0.254)	(0.269)	(0.269)	(0.251)
	$h = 0.06$				$h = 0.08$			
$\phi = 0$, $\psi = 0$	0.897	0.899	0.870	0.792	0.906	0.902	0.884	0.807
	(0.243)	(0.243)	(0.236)	(0.217)	(0.212)	(0.221)	(0.219)	(0.210)
$\phi = 0.2$, $\psi = 0$	0.774	0.779	0.778	0.681	0.774	0.803	0.763	0.688
	(0.237)	(0.245)	(0.245)	(0.244)	(0.213)	(0.223)	(0.226)	(0.221)
$\phi = 0.5$, $\psi = 0$	0.403	0.496	0.522	0.503	0.392	0.502	0.519	0.508
	(0.209)	(0.229)	(0.239)	(0.237)	(0.190)	(0.211)	(0.223)	(0.225)
$\phi = -0.2$, $\psi = 0$	0.960	0.959	0.933	0.875	0.967	0.953	0.941	0.894
	(0.240)	(0.233)	(0.219)	(0.198)	(0.215)	(0.212)	(0.206)	(0.196)
$\phi = -0.5$, $\psi = 0$	0.993	0.996	0.994	0.967	0.994	0.998	0.994	0.962
	(0.217)	(0.205)	(0.185)	(0.166)	(0.197)	(0.190)	(0.180)	(0.176)
$\phi = 0$, $\psi = 0.2$	0.798	0.819	0.785	0.712	0.812	0.817	0.804	0.735
	(0.237)	(0.245)	(0.244)	(0.230)	(0.213)	(0.223)	(0.226)	(0.219)
$\phi = 0$, $\psi = 0.5$	0.670	0.734	0.712	0.657	0.676	0.733	0.726	0.688
	(0.218)	(0.234)	(0.239)	(0.229)	(0.198)	(0.214)	(0.222)	(0.219)

Table 2: Simultaneous coverage probabilities over set G

ARMA(ϕ,ψ)	$\gamma=0$	$\gamma=0.2$	$\gamma=0.4$	$\gamma=0.6$	$\gamma=0$	$\gamma=0.2$	$\gamma=0.4$	$\gamma=0.6$
	$h=0.02$				$h=0.04$			
$\phi=0$, $\psi=0$	0.914	0.919	0.875	0.822	0.914	0.924	0.904	0.836
	(0.477)	(0.465)	(0.430)	(0.367)	(0.368)	(0.346)	(0.346)	(0.306)
$\phi=0.2$, $\psi=0$	0.850	0.857	0.839	0.728	0.838	0.840	0.837	0.745
	(0.463)	(0.462)	(0.439)	(0.387)	(0.358)	(0.366)	(0.356)	(0.327)
$\phi=0.5$, $\psi=0$	0.639	0.694	0.688	0.609	0.584	0.667	0.666	0.616
	(0.392)	(0.412)	(0.408)	(0.376)	(0.311)	(0.337)	(0.343)	(0.330)
$\phi=-0.2$, $\psi=0$	0.943	0.937	0.913	0.841	0.971	0.966	0.948	0.903
	(0.468)	(0.447)	(0.403)	(0.335)	(0.363)	(0.349)	(0.320)	(0.277)
$\phi=-0.5$, $\psi=0$	0.988	0.981	0.968	0.930	0.997	1.000	0.990	0.975
	(0.409)	(0.380)	(0.331)	(0.264)	(0.324)	(0.303)	(0.266)	(0.221)
$\phi=0$, $\psi=0.2$	0.856	0.865	0.833	0.773	0.855	0.864	0.846	0.780
	(0.464)	(0.464)	(0.439)	(0.385)	(0.359)	(0.366)	(0.356)	(0.325)
$\phi=0$, $\psi=0.5$	0.771	0.813	0.779	0.726	0.781	0.809	0.798	0.741
	(0.418)	(0.431)	(0.419)	(0.377)	(0.328)	(0.345)	(0.346)	(0.323)
	$h=0.06$				$h=0.08$			
$\phi=0$, $\psi=0$	0.915	0.926	0.902	0.859	0.913	0.920	0.900	0.856
	(0.314)	(0.315)	(0.306)	(0.281)	(0.283)	(0.288)	(0.285)	(0.272)
$\phi=0.2$, $\psi=0$	0.823	0.846	0.830	0.768	0.816	0.847	0.815	0.756
	(0.307)	(0.318)	(0.316)	(0.300)	(0.277)	(0.290)	(0.294)	(0.286)
$\phi=0.5$, $\psi=0$	0.555	0.643	0.668	0.622	0.529	0.619	0.652	0.610
	(0.271)	(0.297)	(0.309)	(0.305)	(0.248)	(0.274)	(0.289)	(0.293)
$\phi=-0.2$, $\psi=0$	0.969	0.968	0.957	0.901	0.969	0.964	0.946	0.912
	(0.311)	(0.303)	(0.283)	(0.257)	(0.280)	(0.277)	(0.268)	(0.255)
$\phi=-0.5$, $\psi=0$	0.997	0.998	0.993	0.968	0.994	0.997	0.993	0.964
	(0.282)	(0.266)	(0.239)	(0.214)	(0.257)	(0.247)	(0.234)	(0.288)
$\phi=0$, $\psi=0.2$	0.848	0.858	0.843	0.805	0.851	0.853	0.840	0.803
	(0.308)	(0.318)	(0.316)	(0.297)	(0.278)	(0.291)	(0.294)	(0.284)
$\phi=0$, $\psi=0.5$	0.775	0.795	0.808	0.775	0.754	0.792	0.792	0.776
	(0.283)	(0.302)	(0.309)	(0.296)	(0.257)	(0.278)	(0.289)	(0.283)

Table 3: Simultaneous coverage probabilities over set $G.sub$

ARMA(ϕ,ψ)	$\gamma=0$	$\gamma=0.2$	$\gamma=0.4$
	$h=0.04$		
$\phi=0$, $\psi=0$	0.945	0.945	0.937
	(0.555)	(0.546)	(0.511)
$\phi=0.2$, $\psi=0$	0.894	0.903	0.901
	(0.539)	(0.546)	(0.526)
$\phi=-0.2$, $\psi=0$	0.972	0.970	0.962
	(0.545)	(0.520)	(0.475)
$\phi=0$, $\psi=0.2$	0.906	0.914	0.910
	(0.540)	(0.548)	(0.526)

Table 4: Pointwise coverage probabilities, heteroskedastic case

ARMA(ϕ,ψ)	$\gamma = 0$	$\gamma = 0.2$	$\gamma = 0.4$	$\gamma = 0$	$\gamma = 0.2$	$\gamma = 0.4$
	$h = 0.04$					
	G			$G.sub$		
$\phi = 0$, $\psi = 0$	0.896	0.905	0.867	0.921	0.930	0.901
	(0.398)	(0.393)	(0.367)	(0.512)	(0.505)	(0.473)
$\phi = 0.2$, $\psi = 0$	0.795	0.794	0.793	0.842	0.860	0.854
	(0.387)	(0.392)	(0.379)	(0.498)	(0.505)	(0.487)
$\phi = -0.2$, $\psi = 0$	0.966	0.943	0.927	0.971	0.969	0.946
	(0.391)	(0.374)	(0.341)	(0.503)	(0.481)	(0.438)
$\phi = 0$, $\psi = 0.2$	0.817	0.814	0.798	0.852	0.867	0.852
	(0.388)	(0.394)	(0.378)	(0.500)	(0.507)	(0.486)

Table 5: Simultaneous coverage probabilities, heteroskedastic case

ftp://ftp.cpc.ncep.noaa.gov/ndacc/station/jungfrau/hdf/ftir/. It is argued in Franco et al. (2015) that the measurement conditions are very favorable at the Jungfraujoch location due to high dryness and low local pollution. Further details on the ground-based station and on how the measurements are obtained can be found in the aforementioned reference. It is a time series consisting of daily ethane columns (i.e. the number of molecules integrated between the ground and the top of the atmosphere) recorded under clear-sky conditions between September 1994 and August 2014 with a total of 2260 data points. Whenever more than one measurement is taken on one day, a daily mean is considered. The fact that the frequency is daily indicates that the smoothness assumption on the trend function, which we need for nonparametric estimation, should not be cause for concern in this application. Further support that this assumption is satisfied is given by the nature of the data, an instantaneous break is unlikely to occur and changes are expected to be of a gradual form.

The average number of data points per year is 112.6 - giving an indication of the severity of the missing data problem present in this series. This shows that, in line with the above discussion, it is of high importance to use a bootstrap method which can replicate the missing data pattern correctly. In addition, the data exhibit strong seasonality, as ethane degrades faster in summer than it does in winter. This is why the time series displays peaks in winter and is at a low in summer. We take care of this seasonality with the help of Fourier terms. To determine the number of Fourier terms we follow Franco et al. (2015) who argue that adding three Fourier terms capture the intra-annual variability of the data well. We fit the following model, for x_t being the ethane measurements,

$$x_t = \sum_{j=1}^{3} a_j \cos(2j\pi t) + b_j \sin(2j\pi t) + y_t \tag{6.1}$$

and subsequently, fit a nonparametric trend to the residuals from this estimation. The ethane emissions, measured in molecules per cm^2, are displayed together with the seasonal fit in Figure 2.

For the nonparametric estimation, we determine a possible bandwidth using modified cross-validation as described in Section 3.1. In line with the discussion in Chu and Marron (1991) for our series the ordinary leave-one-out cross-validation criterion selects a bandwidth which is too small

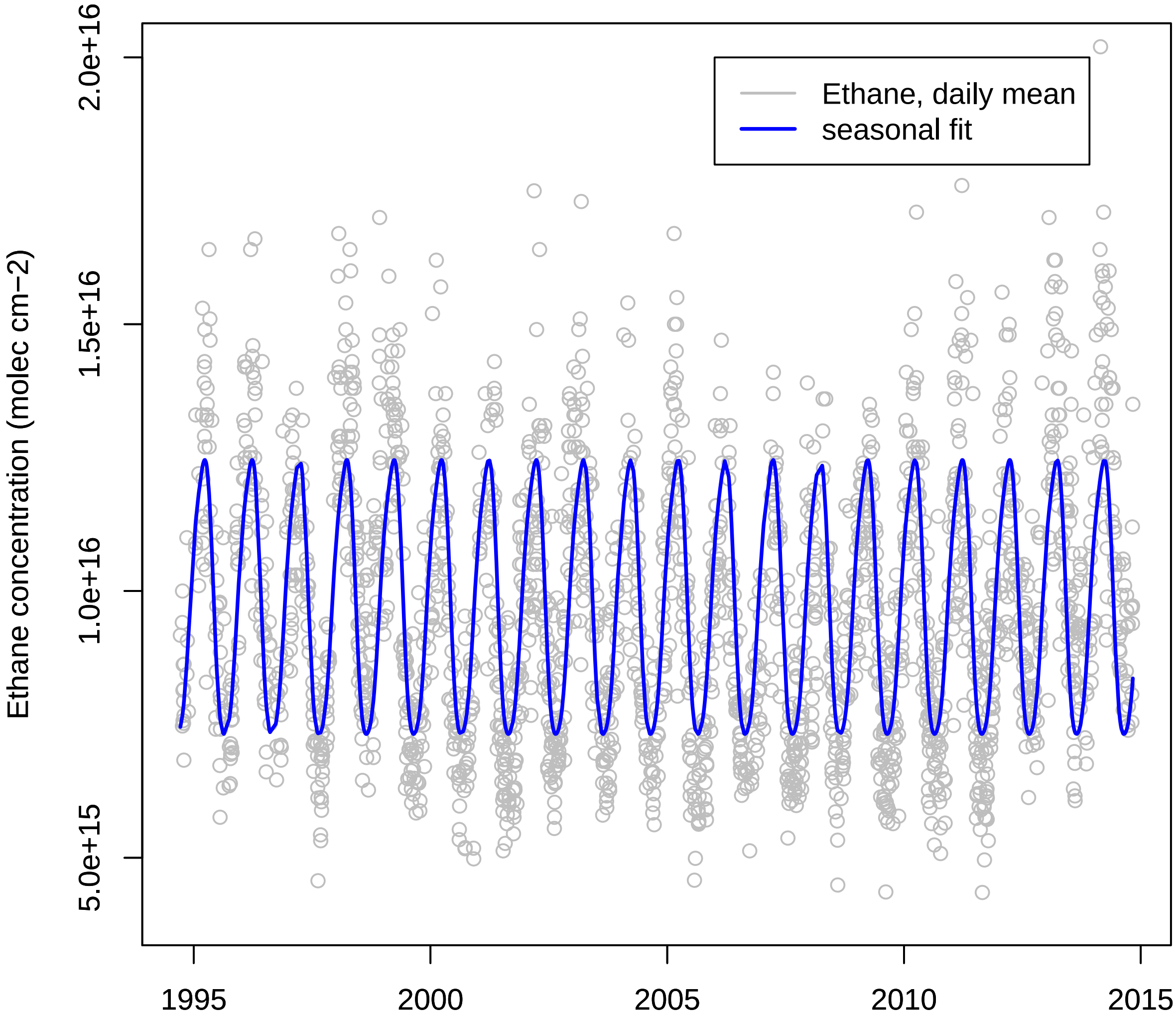


Figure 2: Ethane data with fitted Fourier terms

($h = 0.0006$). This value for the bandwidth parameter gives almost no smoothing of the data and the resulting trend curve would be too rough. Leaving out 5 observations on each side of any point, the modified criterion yields a value of $h = 0.0163$. Albeit being a rather small bandwidth, this value gives a much more reasonable picture of the trend estimate. The resulting nonparametric estimate as well as 95% simultaneous confidence bands are depicted in Figure 3. The confidence bands are simultaneous over the whole sample. Although the validity has not been established, the algorithm works when we cover the whole sample and the results are easier to interpret. The bands are obtained using $B = 999$ replications of the bootstrap procedure and an autoregressive parameter of $\gamma = 0.5$.

When estimating the standard deviation of the residuals after trend estimation by a nonparametric kernel smoother, we see a cyclical pattern with upward trend, similar to the process we generate in our simulations. We plot the estimated standard deviation in Figure 4 which clearly shows that the residuals are heteroskedastic.

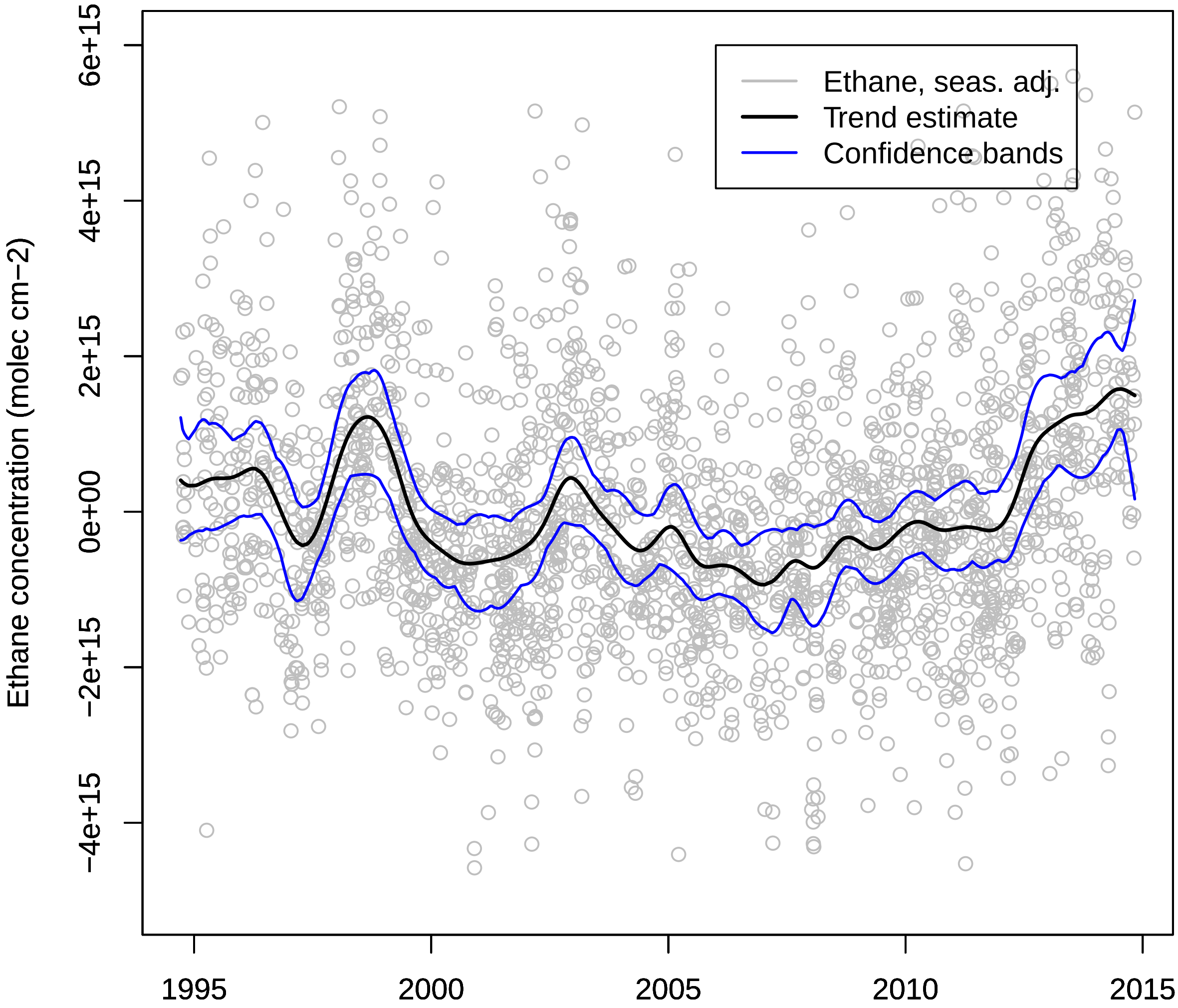


Figure 3: $h = 0.0163$ selected by MCV (Chu and Marron, 1991)

We observe a slight downward trend of the ethane time series until around 2009, with local peaks in 1998 and 2002-2003, and an upward trend thereafter. This general development of the trend supports the findings in Franco et al. (2015) who estimate a linear trend model with a break at the beginning of 2009. They find a negative slope of the trend line before the break and a positive slope after the break. As mentioned by Franco et al. (2015), the initial downward trend can be explained by a general emission reduction since the mid 1980's, of the fossil fuel sources in the Northern Hemisphere. This has also been reported by Simpson et al. (2012). The upward trend seems to be a more recent phenomenon. Studies attribute it to the recent growth in the exploitation of shale gas and tight oil reservoirs, taking place in North America, see e.g. Vinciguerra (2015) and Franco et al. (2016). Since previous studies have mainly used methods based on linear trends, the two local peaks have to our knowledge not yet been analyzed. They can potentially be explained by boreal forest fires which were taking place mainly in Russia during both periods. Geophysical studies have investigated these events in association with anomalies in carbon monoxide emissions (Yurganov et al., 2004, 2005). In such fires, carbon monoxide is co-emitted with ethane, such that these events are likely explanations for the peaks we observe.

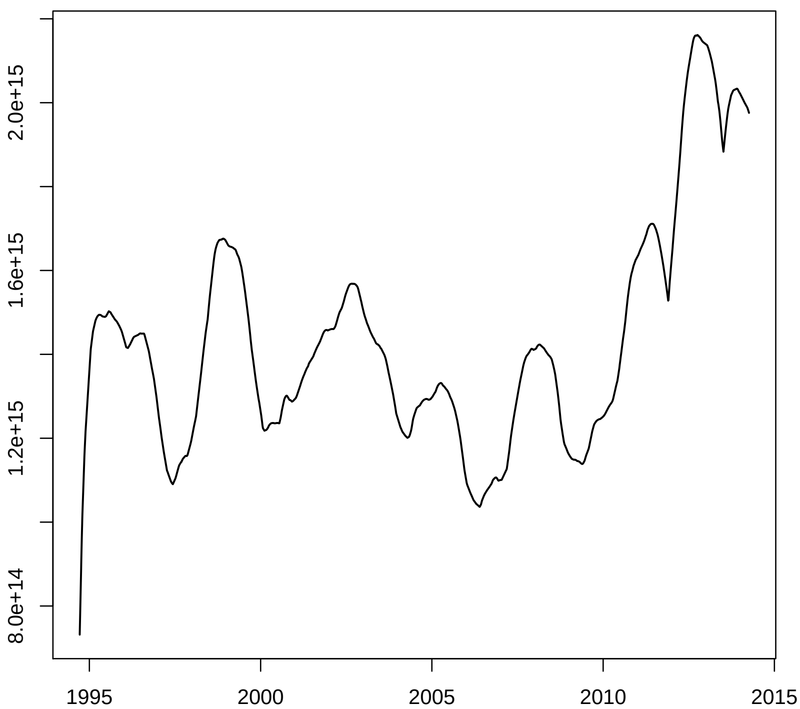


Figure 4: Estimate of the standard deviation of the residuals, obtained using local linear kernel smoother with Epanechnikov kernel and $h = 0.04$

7 Conclusion

In this paper we have proposed a dependent version of the wild bootstrap, the autoregressive wild bootstrap, to construct confidence intervals around a nonparametrically estimated trend. Consistency of the bootstrap has been established such that it can be used to construct pointwise confidence intervals which are shown to provide good coverage in finite samples. In addition, we consider simultaneous confidence bands, constructed using a three-step search algorithm. We have presented a theoretical result on the validity of the bootstrap in this case. The simulation results for this part indicate that strong positive autocorrelation leads to a drop in coverage whenever simultaneous confidence bands are considered. The major advantage of the proposed approach, however, is its broad applicability as it can be used even when the residuals of the model are serially correlated and heteroskedastic. Furthermore, it can be applied without further adjustments when data points are missing. This feature of the autoregressive wild bootstrap is particularly relevant in economic and climatological applications where the problem of missing data is often encountered.

An application to atmospheric ethane measurements from Switzerland demonstrates our methodology. An upward trend in this time series is an indication of increasing atmospheric pollution and it has been visible in the data for the last quarter. This finding is in line with previous studies in geophysics and provides further evidence that an increased activity in shale gas extraction might have caused an increase in the ethane burden measured over the Jungfraujoch. In addition, we find two local peaks in the ethane series, which can be explained by boreal forest fires. Natural limitations of linear trend estimation have prevented these peaks from being discovered in previous research. This underlines the flexibility of our approach compared to parametric methods.

The choice of the autoregressive parameter in the bootstrap reflects a trade-off between variability of the bootstrap sample and capturing the dependence present in the residuals. Although our simulation results suggest that the sensitivity with respect to this parameter is limited, its selection in practice remains an open issue. Theoretical results on the choice of this parameter are not trivial and therefore, left as an exercise for future research.

A Appendix: Proofs of main results

The Appendix is structured as follows. We first state a series of auxiliary lemmas, which will be used in the proofs of the main results and which enhance the structure of these proofs. Second, the proofs of the three main theorems follow. We provide proofs of the auxiliary lemmas in the supplementary appendix.

For the remainder of the proofs we adapt a short-hand notation for the sake of brevity and clarity of certain complex expressions which we will frequently encounter. Let $k_t(\tau) \equiv K\left(\frac{t/n-\tau}{h}\right)$ and $\tilde{k}_t(\tau) \equiv K\left(\frac{t/n-\tau}{\tilde{h}}\right)$. As before, we will use $\sigma_t = \sigma(t/n)$. Further, define $R(k) \equiv Cov(u_t, u_{t+k})$. Given the specifications and assumptions of our model we obtain $R(k) = \sum_{j=0}^{\infty} \psi_j \psi_{j+k}$.

A.1 Auxiliary lemmas

In order to proof the main results of this paper, the following list of lemmas is needed. We refer to all lemmas by its number in the upcoming proofs. The proofs of all lemmas are postponed to part B of this appendix.

Lemma A.1. *Under Assumption 4, we have for any $i \in \mathbb{Z}$ and a constant $C \in \mathbb{R}$*

$$\max_{i\in\mathbb{Z}} \sum_{t=1}^{n} k_t k_{t+i} \leq 2nh \sup_s K(s)^2 \leq Cnh$$
$$\max_{i\in\mathbb{Z}} \sum_{t=1}^{n} k_t^2 k_{t+i}^2 \leq 2nh \sup_s K(s)^4 \leq Cnh.$$

Lemma A.2. *For $|l/n - \tau| < h$ and $|k/n - \tau| < h$, it holds that under Assumption 2*

$$\left|\sigma_k \sigma_l - \sigma(\tau)^2\right| = O(h).$$

Lemma A.3. *Let h satisfy Assumption 5. Under Assumption 4, we have the following limiting expressions of the kernel sums. For $n \to \infty$,*

(i) For $\tau \in [0,1]$:

$$(nh)^{-1} \sum_{t=1}^{n} k_t(\tau) \longrightarrow \int_{\mathbb{R}} K(\omega)\, d\omega$$

(ii) For $\tau \in [0,1]$:

$$(nh)^{-1} \sum_{t=1}^{n} k_t(\tau)^2 \longrightarrow \int_{\mathbb{R}} K(\omega)^2\, d\omega$$

(iii) For $\tau_0 \in [0,1]$ and $\tau, \vartheta \in [-1,1]$:

$$(nh)^{-1} \sum_{t=1}^{n} k_t(\tau_0 + \tau h) k_t(\tau_0 + \vartheta h) \longrightarrow \int_{\mathbb{R}} K(\omega)\, K(\omega + \theta - \tau) d\omega.$$

Lemma A.4. *Given Assumptions 1 to 7, we have that following two bounds*

$$\sup_{\tau\in[0,1]} \mathbb{E}\left(\tilde{m}(t/n) - m(t/n)\right)^2 = O_p\left(\max\left\{\tilde{h}^4, (n\tilde{h})^{-1}\right\}\right),$$

and consequently

$$\frac{1}{n}\sum_{t=1}^{n} \left(\tilde{m}(t/n) - m(t/n)\right)^2 = O_p\left(\max\left\{\tilde{h}^4, (n\tilde{h})^{-1}\right\}\right),$$

where $\tilde{m}(\cdot)$ is defined as in equation 3.2 using bandwidth $\tilde{h}$.

Lemma A.5. *Given Assumption 3 we have that for $R(k) = \sum_{j=0}^{\infty} \psi_j \psi_{j+k}$*

$$\sum_{k=1}^{\infty} k|R(k)| < \infty.$$

Lemma A.6. *Let Assumption 4 hold. For any $k, l \in \mathbb{R}$, we have that*

$$\left|k_t^2(\tau) - k_s^2(\tau)\right| \leq \frac{C|t-s|^2}{(nh)^2}.$$

Lemma A.7. *Let Assumptions 1-6 hold. Then, for all $0 \leq i \leq n-1$,*

(i)

$$\begin{aligned}&\mathbb{E}\left|(nh)^{-1}\sum_{t=1}^{n-i} k_t(\tau)k_{t+i}(\tau)\sigma_t\sigma_{t+i}\left(u_t u_{t+i} - \mathbb{E}\left(u_t u_{t+i}\right)\right)\right| \\ &\quad \leq \beta_i \phi_n + (nh)^{-1/2}\eta_n,\end{aligned}$$

(ii) further, define $t_i = ia + (i-1)b$ with $a(n) = o(n)$ and $b(n) = o(ah)$. Then, for all $-b \leq j \leq b$,

$$\begin{aligned}&\mathbb{E}\left|(nh)^{-1}\sum_{i=1}^{k}\sum_{t=1}^{b-j} k_{t_i+t}(\tau)k_{t_i+t+j}(\tau)\sigma_{t_i+t}\sigma_{t_i+t+j}\left[u_{t_i+t}u_{t_i+t+j} - \mathbb{E}\left(u_{t_i+t}u_{t_i+t+j}\right)\right]\right| \\ &\quad \leq \beta_i \phi_n + (nh)^{-1/2}\eta_n,\end{aligned}$$

where $\sum_{i=1}^{\infty}\beta_i < \infty$, $\lim_{n\to\infty}\phi_n = 0$ and $\limsup_{n\to\infty}\eta_n < \infty$.

Statement (ii) of Lemma A.7 is a variant of (i) which can be used in the blocking technique of Theorem 4.2. The notation here is stated in anticipation of the two theorems, where $a(n)$ is the length of a large block and small blocks are of length $b(n)$.

Lemma A.8. *Under Assumptions 1 to 6, we show that*

(i)

$$\sup_{\tau\in[-1,1]}\left|\mathbb{E}\left(N_{\tau_0,n}(\tau)\right) - B_{as}(\tau_0)\right| = o(1).$$

(ii) and the bootstrap analog:

$$\sup_{\tau\in[-1,1]}\left|\mathbb{E}^*\left(N^*_{\tau_0,n}(\tau)\right) - B_{as}(\tau_0)\right| = o_p(1).$$

A.2 Proofs of theorems

Proof of Theorem 4.1. The main steps of this proof are to first derive the asymptotic bias and variance expression given in the theorem and second, to establish asymptotic normality. The asymptotic bias is unaffected by the presence of heteroskedasticity and therefore, the expression for $B_{as}(\tau)$ is the same as in Bühlmann (1998). The asymptotic variance, however, will be different and we start by deriving its expression.

We look at the centered quantities

$$\hat{m}(\tau) - \mathbb{E}(\hat{m}(\tau)) = \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau)\sigma_t u_t. \tag{A.1}$$

This simplifies further analysis and it is sufficient to work with this quantity in the remainder of this proof, since by Lemma A.8(i) it follows that

$$(nh)^{1/2}\left(\mathbb{E}\left(\hat{m}(\tau)\right)-m(\tau)\right)-B_{as}(\tau)=o(1). \tag{A.2}$$

The asymptotic variance of the estimator is given in the theorem as

$$\sigma_{as}^2(\tau)=\sigma(\tau)^2R(0)\int_{\mathbb{R}}K^2(u)du+2\sigma(\tau)^2\sum_{m=1}^{\infty}R(m)\int_{\mathbb{R}}K^2(u)du\equiv\Gamma_0(\tau)+\Gamma(\tau) \tag{A.3}$$

and can be derived in two main steps. For the two steps, we split up the variance of expression (A.1) into two parts, which will be in line with the two parts of σ_{as}^2 as stated in (A.3)

$$\begin{aligned}(nh)Var\left(\hat{m}(\tau)-\mathbb{E}(\hat{m}(\tau))\right)&=(nh)^{-1}Var\left(\sum_{t=1}^{n}k_t(\tau)\sigma_t u_t\right)=(nh)^{-1}\sum_{t=1}^{n}Var\left(k_t(\tau)\sigma_t u_t\right)\\&\quad+2(nh)^{-1}\sum_{1\le t<s\le n}Cov\left(k_t(\tau)\sigma_t u_t,k_s(\tau)\sigma_s u_s\right)=A_n(\tau)+B_n(\tau)\end{aligned}$$

Splitting up $A_n(\tau)-\Gamma_0(\tau)$ into two parts according to Lemmas A.2 and A.3(ii) and then applying these lemmas to the respective parts, it is easy to show that $|A_n(\tau)-\Gamma_0(\tau)|=o(1)$. We get that

$$\begin{aligned}A_n(\tau)-\Gamma_0(\tau)=&(nh)^{-1}R(0)\sum_{t=1}^{n}k_t^2(\tau)\left(\sigma_t^2-\sigma^2(\tau)\right)\\&+\sigma^2(\tau)R(0)\left((nh)^{-1}\sum_{t=1}^{n}k_t^2(\tau)-\int_{\mathbb{R}}K^2(u)du\right)\\=&A_{1,n}(\tau)+A_{2,n}(\tau),\end{aligned}$$

both $A_{1,n}(\tau)$ and $A_{2,n}(\tau)$ can be shown to be negligible in the limit.

$$\begin{aligned}|A_{1,n}|&\le(nh)^{-1}R(0)\sum_{t=1}^{n}k_t^2(\tau)\left|\sigma_t^2-\sigma^2(\tau)\right|\le R(0)(nh)^{-1}\sum_{t=1}^{n}k_t^2(\tau)Ch=o(1)\\|A_{2,n}|&\le C\left|(nh)^{-1}\sum_{t=1}^{n}k_t^2(\tau)-\int_{\mathbb{R}}K^2(u)du\right|=o(1)\end{aligned}$$

In a similar way, we establish that $|B_n(\tau)-\Gamma(\tau)|=o(1)$. The quantity $B_n(\tau)-\Gamma(\tau)$ can be split into four parts:

$$
\begin{aligned}
B_n(\tau) - \Gamma(\tau) &= 2(nh)^{-1} \sum_{i=1}^{n-1} \sum_{t=1}^{n-i} k_t(\tau) k_{t+i}(\tau) R(i) \left(\sigma_t \sigma_{t+i} - \sigma^2(\tau)\right) \\
&\quad + 2\sigma^2(\tau)(nh)^{-1} \sum_{i=1}^{n-1} R(i) \sum_{t=1}^{n-i} \left(k_t(\tau) k_{t+i}(\tau) - k_t^2(\tau)\right) \\
&\quad + 2\sigma^2(\tau) \sum_{i=1}^{n-1} R(i) \left((nh)^{-1} \sum_{t=1}^{n-i} k_t^2(\tau) - \int_{\mathbb{R}} K^2(u) du\right) \\
&\quad + 2\sigma^2(\tau) \int_{\mathbb{R}} K^2(u) du \left(\sum_{i=1}^{n-1} R(i) - \sum_{j=1}^{\infty} R(j)\right) \\
&= B_{1,n}(\tau) + B_{2,n}(\tau) + B_{3,n}(\tau) + B_{4,n}(\tau)
\end{aligned}
$$

By Lemma A.1 and Lemma A.2 we can show that $B_{1,n}(\tau) = o(1)$

$$
|B_{1,n}(\tau)| \leq 2(nh)^{-1} \sum_{i=1}^{n-1} |R(i)| \, C(nh)h \leq C_1 h \sum_{i=1}^{\infty} |R(i)| \leq C_2 h = o(1).
$$

For the second part, we again use Lemma A.1 and also Lemma A.5:

$$
|B_{2,n}(\tau)| \leq 2\sigma^2(\tau)(nh)^{-2} \sum_{i=1}^{n-1} i|R(i)| \sum_{t=1}^{n-i} |k_t(\tau)| \leq \sigma^2(\tau)(nh)^{-1} \sup_s \{K(s)\} \sum_{i=1}^{n-1} i|R(i)| = o(1)
$$

and hence, $B_{2,n}(\tau) \to 0$ as $n \to \infty$. Next, we establish the asymptotic negligibility of $B_{3,n}(\tau)$

$$
\begin{aligned}
|B_{3,n}(\tau)| &\leq 2\sigma^2(\tau) \sum_{i=1}^{n-1} |R(i)| \left| (nh)^{-1} \sum_{t=1}^{n-i} k_t^2(\tau) - (nh)^{-1} \sum_{t=1}^{n} k_t^2(\tau) \right| \\
&\quad + 2\sigma^2(\tau) \sum_{i=1}^{n-1} |R(i)| \left| (nh)^{-1} \sum_{t=1}^{n} k_t^2(\tau) - \int_{\mathbb{R}} K^2(u) du \right| \\
&\leq 2\sigma^2(\tau) \sum_{i=1}^{n-1} |R(i)| (nh)^{-1} \sum_{t=n-i+1}^{n} k_t^2(\tau) + o(1) \\
&\leq 2\sigma^2(\tau) \sup_s \left\{K(s)^2\right\} \sum_{i=1}^{n-1} i|R(i)| + o(1) \\
&\leq C(nh)^{-1} + o(1),
\end{aligned}
$$

where the $o(1)$ part comes from an application of Lemma A.3. Finally, we look at $B_{4,n}(\tau)$:

$$
|B_{4,n}(\tau)| \leq 2\sigma^2(\tau) \int_{\mathbb{R}} K^2(u) du \left| \sum_{i=1}^{n-1} R(i) - \sum_{j=1}^{\infty} R(j) \right| \leq 2\sigma^2(\tau) \int_{\mathbb{R}} K^2(u) du \sum_{i=n}^{\infty} |R(i)| = o(1).
$$

It follows from the summability of the autocovariances that the tail sum $\sum_{i=n}^{\infty} |R(i)|$ goes to zero as n grows.

Overall, we are left with

$$(nh)\left[Var\left(\hat{m}(\tau)-\mathbb{E}(\hat{m}(\tau))\right)\right]\longrightarrow\sigma(\tau)^2\left[R(0)+2\sum_{m=1}^{\infty}R(m)\right]\int_{\mathbb{R}}K^2(u)du=\Gamma_0(\tau)+\Gamma(\tau),$$

which is the asymptotic variance expression σ^2_{as} as given in the theorem. Now that the asymptotic bias and variance have been derived, we can focus on proving asymptotic normality of $\hat{m}(\tau)-\mathbb{E}(\hat{m}(\tau))=\frac{1}{nh}\sum_{t=1}^{n}k_t(\tau)\sigma_t u_t$. For this, we make use of a blocking argument and then apply the Lindeberg CLT.

Given the MA(∞) representation of u_t we have

$$z_t=\sigma_t u_t=\sigma_t\sum_{j=0}^{\infty}\psi_j\epsilon_{t-j},$$

which we truncate using M such that $M\to\infty$ as $n\to\infty$ to obtain $z_{t,M}\equiv\sigma_t\sum_{j=0}^{M}\psi_j\epsilon_{t-j}$. We first show that the truncation error is negligible such that we only have to consider

$$\bar{z}_{t,M}(\tau)\equiv\frac{1}{nh}\sum_{t=1}^{n}k_t(\tau)z_{t,M}.$$

Denote the truncation error by $\bar{W}_{n,M}=(nh)^{-1}\sum_{t=1}^{n}k_t(\tau)\sigma_t\sum_{j=M+1}^{\infty}\psi_j\epsilon_{t-j}$. It is easy to see that $\mathbb{E}\left(\bar{W}_{n,M}\right)=0$ so it remains to show that $Var\left(\bar{W}_{n,M}\right)=o(1)$. Let $R_W(k)=\sum_{j=M+1}^{\infty}\psi_j\psi_{j+|k|}$.

$$\begin{aligned}
\mathbb{E}\left(\bar{W}_{n,M}\right)^2&=\frac{1}{(nh)^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\sigma_t\sigma_s k_t(\tau)k_s(\tau)R_W(t-s)\\
&=\frac{1}{(nh)^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\left[\sigma_t\sigma_s-\sigma(\tau)^2+\sigma(\tau)^2\right]k_t(\tau)k_s(\tau)R_W(t-s)\\
&\leq\frac{1}{(nh)^2}\sigma(\tau)^2\sum_{t=1}^{n}\sum_{s=1}^{n}k_t(\tau)k_s(\tau)R_W(t-s)+O(h)\\
&\leq\sigma(\tau)^2\sum_{k=-n+1}^{n-1}R_W(k)\left[\frac{1}{nh}\sum_{t=1}^{n}k_t(\tau)\right]^2+O(h)\\
&\leq\sigma(\tau)^2\sum_{k=-n+1}^{n-1}\sum_{j=M+1}^{\infty}\psi_j\psi_{j+|k|}+O(h)=\sigma(\tau)^2\sum_{j=M+1}^{\infty}\psi_j\sum_{k=-n+1}^{n-1}\psi_{j+|k|}+O(h)\\
&\leq C\sum_{j=M+1}^{\infty}|\psi_j|+O(h)\leq\frac{C}{M}\sum_{j=M+1}^{\infty}j|\psi_j|+O(h)=o(1),
\end{aligned}$$

where the first inequality follows by the same reasoning as before, which explains why the first term is $o(1)$. Then we use the fact that $\frac{1}{nh}\sum_{t=1}^{n}k_t(\tau)\to 1$ and that $\frac{j}{M}>1$. The final step makes use of the summability condition and the fact that $M\to\infty$.

We next split $\bar{z}_{t,M}(\tau)$ into two types of blocks: small, negligible blocks $Y_{n,i}(\tau)$ and dominating blocks $X_{n,i}(\tau)$. Define

$$X_{n,i}(\tau) = \frac{1}{nh} \sum_{t=(i-1)(a+b)+1}^{ia+(i-1)b} k_t(\tau) z_{t,M},$$

$$Y_{n,i}(\tau) = \frac{1}{nh} \sum_{t=ia+(i-1)b+1}^{i(a+b)} k_t(\tau) z_{t,M},$$

so that $\bar{z}_{t,M}$ splits up into k blocks with $k = \left\lfloor \frac{n}{a+b} \right\rfloor$

$$\bar{z}_{t,M}(\tau) = \sum_{i=1}^{k} X_{n,i}(\tau) + \sum_{i=1}^{k} Y_{n,i}(\tau).$$

We need the block length to be increasing with n under some conditions. Specifically, let $a(n) = o(n)$ and $b(n) = o(ah)$ while $b(n)/M \to \infty$ as $n \to \infty$.

We next show that the small blocks are asymptotically negligible:

$$\sqrt{nh} \sum_{i=1}^{k} Y_{n,i}(\tau) = o_p(1). \tag{A.4}$$

It is easy to see that $\mathbb{E}\left(\sqrt{nh}\sum_{i=1}^{k} Y_{n,i}(\tau)\right) = 0$. Further, we have that the $z_{t,M}$'s are M-dependent. Therefore, consider n large enough such that $a(n) > M$. Then, letting $R_M(k) = \sum_{j=0}^{M} \psi_j \psi_{j+k}$ and $\mathcal{Y}_i \equiv \{t : ia + (i-1)b + 1 \leq t \leq i(a+b)\}$,

$$\begin{aligned}
Var\left(\sqrt{nh}\sum_{i=1}^{k} Y_{n,i}(\tau)\right) &= (nh)^{-1} \sum_{i=1}^{k} Var(Y_{n,i}(\tau)) = (nh)^{-1} \sum_{i=1}^{k} \sum_{t \in \mathcal{Y}_i} \sum_{s \in \mathcal{Y}_i} \sigma_t \sigma_s k_t(\tau) k_s(\tau) R_M(t-s) \\
&\leq (nh)^{-1} k \sum_{j=-b}^{b} R_M(j) \sum_{t=0}^{b-|j|} \left[\sigma(\tau)^2 + Ch\right] k_{a+t}(\tau) k_{a+t+j}(\tau) \\
&= O\left(\frac{b}{ah}\right) + O\left(\frac{b}{a}\right).
\end{aligned}$$

To obtain the first inequality, we use a similar argument is used in the proof of Lemma A.2 together with the fact that the Kernel function is 0, whenever $|t/n - \tau| < h$ or $|s/n - \tau| < h$ is violated. In the last step we use the bound $(nh)^{-1} \sum_{t=a}^{a+b} k_t(\tau) = O\left(b/(nh)\right)$. As our assumptions imply that $O\left(\frac{b}{ah}\right) = o(1)$, this completes the proof of A.4.

The final step is to show that large blocks follow a CLT and that the asymptotic variance coincides with the one given in the theorem. Using the Lindeberg CLT (see, e.g., Davidson (2002), Theorem 23.6), we can show that

$$\sqrt{nh} \sum_{i=1}^{k} X_{n,i}(\tau) \xrightarrow{d} \mathcal{N}(0, \sigma_{as}^2(\tau)). \tag{A.5}$$

Define $\mathcal{X}_i \equiv \{t : (i-1)(a+b) + 1 \leq t \leq ia + (i-1)b\}$. Similarly to the previous step, we have that

$\mathbb{E}\left(\sqrt{nh}\sum_{i=1}^{k} X_{n,i}(\tau)\right) = 0$. Consider, again, n sufficiently large such that by M-dependence, big blocks are independent. Specifically, let n such that $b(n) > M$.

$$
\begin{aligned}
Var\left(\sqrt{nh}\sum_{i=1}^{k} X_{n,i}(\tau)\right) &= (nh)^{-1}\sum_{i=1}^{k}\mathbb{E}(X_{n,i}(\tau))^2 = (nh)^{-1}\sum_{i=1}^{k}\sum_{t\in\mathcal{X}_i} k_t(\tau)^2\sigma_t^2 R(0) \\
&+ 2(nh)^{-1}\sum_{i=1}^{k}\sum_{t,s\in\mathcal{X}_i;t<s} k_t(\tau)k_s(\tau)\sigma_t\sigma_s R(t-s) = D_n(\tau) + E_n(\tau)
\end{aligned}
$$

The derivation is in analogy to the derivation of the asymptotic variance above. Hence, we first show that $|D_n(\tau) - \Gamma_0(\tau)| = o(1)$ and second that $|E_n(\tau) - \Gamma(\tau)| = o(1)$. In the remainder, $R_M(0)$ denotes the truncated version of $R(0)$ and is defined as $R_M(0) = \sum_{j=0}^{M}\psi_j^2$. So, we obtain for the first part:

$$
\begin{aligned}
D_n(\tau) - \Gamma_0(\tau) =& (nh)^{-1} R_M(0)\sum_{i=1}^{k}\sum_{t\in\mathcal{X}_i} k_t^2(\tau)\left(\sigma_t^2 - \sigma^2(\tau)\right) \\
&+ \sigma^2(\tau)R_M(0)\left((nh)^{-1}k\sum_{t=1}^{a} k_t^2(\tau) - \int_{\mathbb{R}} K^2(u)du\right) \\
&+ \sigma^2(\tau)\int_{\mathbb{R}} K^2(u)du\,(R_M(0) - R(0)) \\
=& D_{1,n}(\tau) + D_{2,n}(\tau) + D_{3,n}(\tau),
\end{aligned}
$$

The negligibility of the first part follows the same reasoning as for part $A_{1,n}(\tau)$ in the asymptotic variance derivation. Therefore, $D_{1,n}(\tau) \to 0$ as $n \to \infty$. Since $k/n = 1/(a+b)$ is of order $1/a$, the integral approximation holds and we get correspondence between $D_{2,n}(\tau)$ and $A_{2,n}(\tau)$ above. For the third part, clearly, $R_M(0) \to R(0)$, because $M(n) \to \infty$. Thus, we have established that $|D_n(\tau) - \Gamma_0(\tau)| = o(1)$ holds. For the covariance part, we proceed in analogy to part $B_n(\tau)$ above and get the following four terms:

$$
\begin{aligned}
E_n(\tau) - \Gamma(\tau) &= 2(nh)^{-1}\sum_{j=1}^{k}\sum_{i,t\in\mathcal{X}_j,t<i} k_t(\tau)k_{t+i}(\tau)R_M(i)\left(\sigma_t\sigma_{t+i} - \sigma^2(\tau)\right) \\
&+ 2\sigma^2(\tau)k(nh)^{-1}\sum_{i=1}^{a-1} R_M(i)\sum_{t=1}^{a-i}\left(k_t(\tau)k_{t+i}(\tau) - k_t^2(\tau)\right) \\
&+ 2\sigma^2(\tau)\sum_{i=1}^{a-1} R_M(i)\left((nh)^{-1}k\sum_{t=1}^{a-i} k_t^2(\tau) - \int_{\mathbb{R}} K^2(u)du\right) \\
&+ 2\sigma^2(\tau)\int_{\mathbb{R}} K^2(u)du\left(\sum_{i=1}^{a-1} R_M(i) - \sum_{j=1}^{\infty} R(j)\right) \\
&= E_{1,n}(\tau) + E_{2,n}(\tau) + E_{3,n}(\tau) + E_{4,n}(\tau)
\end{aligned}
$$

The negligibility of all four parts follow in a similar way than with the four parts of $B_n(\tau)$ above, keeping in mind that $a(n) \to \infty$ and $M(n) \to \infty$ as $n \to \infty$.

Overall, we have that

$$Var\left(\sqrt{nh}\sum_{i=1}^{k}X_{n,i}(\tau)\right)\longrightarrow\sigma(\tau)^2\left[R(0)+2\sum_{m=1}^{\infty}R(m)\right]\int_{\mathbb{R}}K^2(u)du=\Gamma_0(\tau)+\Gamma(\tau),$$

which corresponds with $\sigma_{as}^2(\tau)$. The final step in order to show that (A.5) holds is to verify the Lindeberg condition. For this we need to show that, for $\kappa>0$,

$$\sum_{i=1}^{k}\mathbb{E}\left[\frac{X_{n,i}(\tau)^2}{\omega_n^2}\mathbb{1}_{\{|X_{n,i}(\tau)/\omega_n|>\kappa\}}\right]=o(1) \tag{A.6}$$

with $\omega_n^2=Var\left(\sum_{i=1}^{k}X_{n,i}\right)=O(nh)$. We have that

$$\sum_{i=1}^{k}\mathbb{E}\left[\frac{X_{n,i}(\tau)^2}{\omega_n^2}\mathbb{1}_{\{|X_{n,i}(\tau)/\omega_n|>\kappa\}}\right]=\sum_{i=1}^{k}\int_{\mathbb{R}}\frac{X_{n,i}(\tau)^2}{\omega_n^2}\mathbb{1}_{\{|X_{n,i}(\tau)/\omega_n|>\kappa\}}d\mathbb{P}$$
$$\leq\sum_{i=1}^{k}\int_{\mathbb{R}}\frac{X_{n,i}(\tau)^2}{\omega_n^2}\left[\frac{X_{n,i}(\tau)^2}{\omega_n^2\kappa^2}\right]\mathbb{1}_{\{|X_{n,i}(\tau)/\omega_n|>\kappa\}}d\mathbb{P}\leq\frac{1}{\kappa^2}\sum_{i=1}^{k}\int_{\mathbb{R}}\frac{X_{n,i}(\tau)^4}{\omega_n^4}d\mathbb{P}=\frac{1}{\kappa^2}\sum_{i=1}^{k}\omega_n^{-4}\,\mathbb{E}\left|X_{n,i}(\tau)\right|^4.$$

Consider the fourth moment

$$\mathbb{E}\left|X_{n,i}(\tau)\right|^4=\mathbb{E}\left|(nh)^{-1}\sum_{t\in\mathcal{X}_i}k_tz_{t,M}\right|^4=(nh)^{-4}\,\mathbb{E}\left|\sum_{t\in\mathcal{X}_i}k_t(\tau)z_{t,M}\right|^4\leq\frac{a^3}{(nh)^4}\sum_{t\in\mathcal{X}_i}\mathbb{E}\left|k_t(\tau)z_{t,M}\right|^4$$
$$=a^3\sup_t\left\{\mathbb{E}\left|z_{t,M}\right|^4\right\}(nh)^{-4}\sum_{t\in\mathcal{X}_i}k_t^4(\tau),$$

where we use the c_r-inequality (see Davidson (2002), p.140). Further, $(nh)^{-4}\sum_{t\in\mathcal{X}_i}k_t^4(\tau)=O\left(\frac{a}{(nh)^4}\right)$. For $\sup_t\left\{\mathbb{E}\left|z_{t,M}\right|^4\right\}$ we have that, using the stationarity of ϵ_t,

$$\sup_t\left\{\mathbb{E}\left|z_{t,M}\right|^4\right\}=\sup_t\left\{\mathbb{E}\left|\sigma_t\sum_{j=0}^{M}\psi_j\epsilon_{t-j}\right|^4\right\}\leq\left(\sup_t\left\{\sigma_t\right\}\sum_{j=0}^{M}\left(\mathbb{E}\left|\psi_j\epsilon_{t-j}\right|^4\right)^{1/4}\right)^4$$
$$=\left(\sup_t\left\{\sigma_t\right\}\sum_{j=0}^{M}\psi_j\left(\mathbb{E}\left|\epsilon_{t-j}\right|^4\right)^{1/4}\right)^4=\left[\sup_t\left\{\sigma_t\right\}\right]^4\mathbb{E}\left|\epsilon_t\right|^4\left(\sum_{j=0}^{M}\psi_j\right)^4=O(1).$$

The second step uses Minkowski's inequality. Overall, (A.6) is satisfied. To see this we use, again, that $a(n)=o(n)$ and $ka=O(n)$,

$$\sum_{i=1}^{k}\mathbb{E}\left[\frac{X_{n,i}(\tau)^2}{\sigma_n^2}\mathbb{1}_{\{|X_{n,i}(\tau)/\sigma_n|>\kappa\}}\right]\leq\kappa^{-2}\sum_{i=1}^{k}\omega_n^{-4}a^3(nh)^{-4}\sum_{t\in\mathcal{X}_i}k_t^4(\tau)\,\mathbb{E}\left|z_{t,M}\right|^4$$
$$=O\left(\frac{a^3}{n^5h^2}\right)=o(1).$$

□

Proof of Theorem 4.2. The proof of this theorem follows the same basic structure as the proof of Theorem 4.1. The main steps resemble the steps made in the above proof, the difference is that we consider bootstrap quantities and thus, conduct most part of the analysis conditional on the data. We first derive the asymptotic variance expression of the estimator in the bootstrap world and then establish asymptotic normality. We will then observe that the asymptotic distribution is the same as the one of the estimator derived in Theorem 4.1, which completes the proof. In analogy to above, we consider centered quantities

$$\hat{m}^*(\tau) - \mathbb{E}(\hat{m}^*(\tau)) = \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau) z_t^* = \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau)\xi_t^* \hat{z}_t. \tag{A.7}$$

Recall that $z_t^* = \xi_t^* \hat{z}_t$ and $\xi_t^* = \gamma\xi_{t-1}^* + \nu_t^*$ with $\nu_1^*, ..., \nu_t^* \sim \mathcal{N}\left(0, 1-\gamma^2\right)$. The asymptotic bias is the same as in the non-bootstrap case. To see that centering is a valid step, we note that the bootstrap analogue of (A.2) immediately follows by an application of Lemma A.8(ii). Specifically,

$$(nh)^{1/2}\left(\mathbb{E}^*\left(\hat{m}^*(\tau)\right) - \tilde{m}(\tau)\right) - B_{as}(\tau) = o_p(1). \tag{A.8}$$

We start out by deriving the asymptotic variance of the bootstrap estimate.

$$\begin{aligned}(nh)Var^*\left(\hat{m}^*(\tau) - \mathbb{E}(\hat{m}^*(\tau))\right) = (nh)^{-1}Var^*\left(\sum_{t=1}^{n} k_t(\tau)\xi_t^*\hat{z}_t\right) = (nh)^{-1}\sum_{t=1}^{n} Var^*\left(k_t(\tau)\xi_t^*\hat{z}_t\right)\\ + 2(nh)^{-1}\sum_{1\le k<l\le n} Cov^*\left(k_k(\tau)\xi_k^*\hat{z}_k, k_l(\tau)\xi_l^*\hat{z}_l\right) = A_n^*(\tau) + B_n^*(\tau)\end{aligned}$$

The two terms are the bootstrap versions of $A_n(\tau)$ and $B_n(\tau)$ in the proof of Theorem 4.1. We will show that the limits as $n \to \infty$ of the two bootstrap expressions are the same as the ones of the corresponding non-bootstrap quantities. $\Gamma_0(\tau)$ will be the limit of $A_n^*(\tau)$ and $\Gamma(\tau)$ the limit of $B_n^*(\tau)$. We start by further investigating $A_n^*(\tau)$.

$$\begin{aligned}A_n^*(\tau) &= (nh)^{-1}\sum_{t=1}^{n} k_t^2(\tau)\hat{z}_t^2 Var^*(\xi_t^*)\\ &= (nh)^{-1}\sum_{t=1}^{n} k_t^2(\tau)\sigma_t^2 u_t^2 + (nh)^{-1}\sum_{t=1}^{n} k_t^2(\tau)\left(m(t/n) - \tilde{m}(t/n)\right)^2\\ &\quad + (nh)^{-1}\sum_{t=1}^{n} k_t^2(\tau)\sigma_t u_t\left(m(t/n) - \tilde{m}(t/n)\right)\\ &= A_{1,n}^*(\tau) + A_{2,n}^*(\tau) + A_{3,n}^*(\tau),\end{aligned} \tag{A.9}$$

where we use the fact that $Var^*(\xi_t^*) = 1$. Now, we can show that the last two parts are asymptotically negligible. First, we look at $A^*_{2,n}(\tau)$ and apply Lemma A.4:

$$\begin{aligned} A^*_{2,n}(\tau) &= (nh)^{-1}\sum_{t=1}^{n} k_t^2(\tau)\left(m(t/n) - \tilde{m}(t/n)\right)^2 \\ &\leq h^{-1}\sup_s \left\{K(s)^2\right\} n^{-1}\sum_{t=1}^{n}\left(m(t/n) - \tilde{m}(t/n)\right)^2 \\ &= O_p\left(\max\left\{\tilde{h}^4/h, (n\tilde{h})^{-1}/h\right\}\right) \end{aligned}$$

and hence, $A^*_{2,n}(\tau) \to_p 0$ as $n \to 0$. Second, we bound $A^*_{3,n}(\tau)$ making use of the Cauchy-Schwarz inequality in combination with Lemma A.4

$$\begin{aligned} \left[A^*_{3,n}(\tau)\right]^2 &= \left[(nh)^{-1}\sum_{t=1}^{n} k_t^2(\tau)\sigma_t u_t\left(m(t/n) - \tilde{m}(t/n)\right)\right]^2 \\ &\leq \left[(nh)^{-1}\sum_{t=1}^{n} k_t^2(\tau)\sigma_t^2 u_t^2\right]\left[(nh)^{-1}\sum_{t=1}^{n}\left(m(t/n) - \tilde{m}(t/n)\right)^2\right] \\ &= O_p\left(\max\left\{\tilde{h}^4/h, (n\tilde{h})^{-1}/h\right\}\right) \end{aligned}$$

such that also for the final part it holds that $A^*_{3,n}(\tau) \to_p 0$ as $n \to 0$. We are now going to show that the dominating part of $(nh)A^*_n(\tau)$, which is $A^*_{1,n}(\tau)$, is tightly linked to a long-run variance estimator for linear processes. Therefore, Lemma A.7 will play a big role in the remainder of the proof. This also explains the choice of notation.

$$\hat{\Gamma}_0(\tau) \equiv A^*_{1,n}(\tau) = (nh)^{-1}\sum_{t=1}^{n} k_t^2(\tau)\sigma_t^2 u_t^2 \to_p \sigma(\tau)^2 R(0)\int_{\mathbb{R}} K^2(u)du = \Gamma_0(\tau). \tag{A.10}$$

By the triangle inequality we have that

$$\left|\hat{\Gamma}_0(\tau) - \Gamma_0(\tau)\right| \leq \left|\hat{\Gamma}_0(\tau) - \mathbb{E}\,\hat{\Gamma}_0(\tau)\right| + \left|\mathbb{E}\,\hat{\Gamma}_0(\tau) - \Gamma_0(\tau)\right|.$$

We will now first show that for all $\tau \in (0,1)$

$$\hat{\Gamma}_0(\tau) - \mathbb{E}\,\hat{\Gamma}_0(\tau) = (nh)^{-1}\sum_{t=1}^{n} k_t^2(\tau)\sigma_t^2\left[u_t^2 - \mathbb{E}(u_t^2)\right] = o_p(1). \tag{A.11}$$

It follows from Markov's inequality and Lemma A.7(i) for the case $i = 0$ that

$$\mathbb{P}\left(\left|\hat{\Gamma}_0(\tau) - \mathbb{E}\,\hat{\Gamma}_0(\tau)\right| > \delta\right) \leq \delta^{-1}\,\mathbb{E}\left|(nh)^{-1}\sum_{t=1}^{n} k_t^2(\tau)\sigma_t^2\left[u_t^2 - \mathbb{E}(u_t^2)\right]\right| \leq \delta^{-1}\left[\beta_0\phi_n + (nh)^{-1/2}\eta_n\right].$$

Since $\phi_n \to 0$ as $n \to \infty$ and $\limsup_{n\to\infty}\eta_n < \infty$, we have that the probability goes to zero overall as $n \to \infty$. This proves (A.11). Next we show that

$$\mathbb{E}\,\hat{\Gamma}_0(\tau) - \Gamma_0(\tau) = (nh)^{-1}\sum_{t=1}^{n} k_t^2(\tau)\sigma_t^2 R(0) - \sigma(\tau)^2 R(0)\int_{\mathbb{R}} K^2(u)du = o(1), \tag{A.12}$$

by splitting the expression up into two parts as follows

$$\mathbb{E}\,\hat{\Gamma}_0(\tau) - \Gamma_0(\tau) = (nh)^{-1} R(0) \sum_{t=1}^{n} k_t^2(\tau) \left(\sigma_t^2 - \sigma(\tau)^2\right) \\ + R(0)\sigma(\tau)^2 \left((nh)^{-1} \sum_{t=1}^{n} k_t^2(\tau) - \int_{\mathbb{R}} K^2(u) du \right).$$

By Lemmas A.2 and A.3 it immediately follows that both parts go to zero as $n \to \infty$. Together this is sufficient to establish (A.12) and, in turn, from (A.11) and (A.12) we can conclude that (A.10) holds.

We finish the derivation of the asymptotic variance by investigating the covariance term $B_n^*(\tau)$. Again, we establish a link to long-run variance estimation in the case of linear processes. First, it is easy to show that $Cov^*\left(\xi_k^*, \xi_l^*\right) = \gamma^{|k-l|} = \theta^{\frac{|k-l|}{\ell}}$. Using this result we obtain

$$\begin{aligned} B_n^*(\tau) &= 2(nh)^{-1} \sum_{1 \le k < l \le n} Cov^*\left(k_k(\tau)\xi_k^* \hat{z}_k, k_l(\tau)\xi_l^* \hat{z}_l\right) \\ &= 2(nh)^{-1} \sum_{1 \le k < l \le n} k_k(\tau) k_l(\tau) \sigma_k \sigma_l u_k u_l \theta^{\frac{l-k}{\ell}} \\ &\quad + 2(nh)^{-1} \sum_{1 \le k < l \le n} k_k(\tau) k_l(\tau) \left[m(k/n) - \tilde{m}(k/n)\right] \sigma_l u_l \theta^{\frac{l-k}{\ell}} \\ &\quad + 2(nh)^{-1} \sum_{1 \le k < l \le n} k_k(\tau) k_l(\tau) \left[m(l/n) - \tilde{m}(l/n)\right] \sigma_k u_k \theta^{\frac{l-k}{\ell}} \\ &\quad + 2(nh)^{-1} \sum_{1 \le k < l \le n} k_k(\tau) k_l(\tau) \left[m(k/n) - \tilde{m}(k/n)\right] \left[m(l/n) - \tilde{m}(l/n)\right] \theta^{\frac{l-k}{\ell}} \\ &= B_{1,n}^*(\tau) + B_{2,n}^*(\tau) + B_{3,n}^*(\tau) + B_{4,n}^*(\tau), \end{aligned}$$

where the four parts are emerging from replacing $\hat{z}_k \hat{z}_l$ in the third step by

$$\begin{aligned} \hat{z}_k \hat{z}_l &= \left[y_k - \tilde{m}(k/n)\right] \left[y_l - \tilde{m}(l/n)\right] \\ &= \sigma_k \sigma_l u_k u_l + \left[m(k/n) - \tilde{m}(k/n)\right] \sigma_l u_l + \left[m(l/n) - \tilde{m}(l/n)\right] \sigma_k u_k \\ &\quad + \left[m(k/n) - \tilde{m}(k/n)\right] \left[m(l/n) - \tilde{m}(l/n)\right]. \end{aligned}$$

We show that the last three parts are asymptotically negligible and that for $B_{1,n}^*(\tau)$ it holds that

$$\hat{\Gamma}(\tau) \equiv 2(nh)^{-1} \sum_{k=1}^{n-1} \sum_{l=k+1}^{n} k_k(\tau) k_l(\tau) \sigma_k \sigma_l \theta^{\frac{l-k}{\ell}} u_k u_l \to_p 2\sigma(\tau)^2 \int_{\mathbb{R}} K(u)^2 du \sum_{j=1}^{\infty} R(j) = \Gamma(\tau). \quad \text{(A.13)}$$

We treat $B_{2,n}^*(\tau)$ first. We use Markov's inequality first:

$$\mathbb{P}\left(\left|B_{2,n}^*(\tau)\right| \ge a\right) \le \frac{\mathbb{E}\left|B_{2,n}^*(\tau)\right|}{a}$$

Then, by the Cauchy-Schwartz inequality and Lemmas A.1 and A.4 it follows:

$$
\begin{aligned}
\mathbb{E}\left|B_{2,n}^*(\tau)\right| &= \mathbb{E}\left|2(nh)^{-1}\sum_{i=1}^{n}\theta^{\frac{i}{\ell}}\sum_{t=1}^{n-i}k_t(\tau)k_{t+i}(\tau)\left[m((t+i)/n)-\tilde{m}((t+i)/n)\right]\sigma_t u_t\right| \\
&\le 2(nh)^{-1}\sum_{i=1}^{n}\theta^{\frac{i}{\ell}}\sum_{t=1}^{n-i}k_t^2(\tau)k_{t+i}^2(\tau)\left(\mathbb{E}\left[m((t+i)/n)-\tilde{m}((t+i)/n)\right]^2\right)^{1/2}\sigma_t\left(\mathbb{E}\,u_t^2\right)^{1/2} \\
&\le 2\max_t\sigma_t(\mathbb{E}\,u_t^2)^{1/2}(nh)^{-1}\sup_\tau\mathbb{E}\left[m(\tau)-\tilde{m}(\tau)\right]^2)^{1/2}\sum_{i=1}^{n}\theta^{\frac{i}{\ell}}\sum_{t=1}^{n-i}k_t^2(\tau)k_{t+i}^2(\tau) \\
&\le 2C(\sup_\tau\mathbb{E}\left[m(\tau)-\tilde{m}(\tau)\right]^2)^{1/2}\max_{0\le i\le n}(nh)^{-1}\sum_{t=1}^{n}k_t^2(\tau)k_{t+i}^2(\tau)\frac{1}{1-\theta^{1/\ell}} \\
&= O\left(\max\left\{\tilde{h}^2,(n\tilde{h})^{-1/2}\right\}\right)O(1)o(\ell).
\end{aligned}
$$

Hence, by Assumption 6, $\mathbb{E}\left|B_{2,n}^*(\tau)\right| = o(1)$. Now, $B_{3,n}^*(\tau)$ and $B_{4,n}^*(\tau)$ can be treated analogously.

To proceed, we again make a change of variables as in Jansson (2002) and write

$$
\hat{\Gamma}(\tau) = 2(nh)^{-1}\sum_{i=1}^{n-1}\theta^{i/\ell}\sum_{t=1}^{n-i}k_t(\tau)k_{t+i}(\tau)\sigma_t\sigma_{t+i}u_tu_{t+i}.
$$

By the triangle inequality we have that

$$
\left|\hat{\Gamma}(\tau)-\Gamma(\tau)\right| \le \left|\hat{\Gamma}(\tau)-\mathbb{E}\,\hat{\Gamma}(\tau)\right| + \left|\mathbb{E}\,\hat{\Gamma}(\tau)-\Gamma(\tau)\right|.
$$

We will now first show that for all $\tau\in(0,1)$

$$
\hat{\Gamma}(\tau)-\mathbb{E}\,\hat{\Gamma}(\tau) = 2(nh)^{-1}\sum_{i=1}^{n-1}\theta^{i/\ell}\sum_{t=1}^{n-i}k_t(\tau)k_{t+i}(\tau)\sigma_t\sigma_{t+i}\left(u_tu_{t+i}-\mathbb{E}\,u_tu_{t+i}\right) = o_p(1). \tag{A.14}
$$

It follows from Markov's inequality and Lemma A.7(i) that

$$
\begin{aligned}
\mathbb{P}\left(\left|\hat{\Gamma}(\tau)-\mathbb{E}\,\hat{\Gamma}(\tau)\right|>\delta\right) &\le \delta^{-1}\sum_{i=1}^{n-1}\theta^{i/\ell}\,\mathbb{E}\left|(nh)^{-1}\sum_{t=1}^{n-i}k_t(\tau)k_{t+i}(\tau)\sigma_t\sigma_{t+i}\left(u_tu_{t+i}-\mathbb{E}\,u_tu_{t+i}\right)\right| \\
&\le \delta^{-1}\sum_{i=1}^{n-1}\theta^{i/\ell}[\beta_i\phi_n+(nh)^{-1/2}\eta_n] \le \delta^{-1}\phi_n\sum_{i=1}^{n-1}\beta_i+(nh)^{-1/2}\delta^{-1}\eta_n\frac{1-\theta^{n/\ell}}{1-\theta^{1/\ell}}.
\end{aligned}
$$

As $\sum_{i=0}^{\infty}\beta_i<\infty$ and $\phi_n\to 0$ as $n\to\infty$, the first part converges to zero as $n\to\infty$. For the second part, note that $\ell^{-1}\frac{1-\theta^{n/\ell}}{1-\theta^{1/\ell}} = -\frac{1}{\ln\theta}+o(1)$. As by Assumption 6, $\ell(nh)^{-1/2}=o(1)$, the second part also converges to zero as $n\to\infty$. This completes the proof of (A.14).

Next we show that

$$
\begin{aligned}
\mathbb{E}\,\hat{\Gamma}(\tau)-\Gamma(\tau) = {} & 2(nh)^{-1}\sum_{i=1}^{n-1}\theta^{i/\ell}\sum_{t=1}^{n-i}k_t(\tau)k_{t+i}(\tau)\sigma_t\sigma_{t+i}R(i) \\
& - 2\sigma(\tau)^2\int_{\mathbb{R}}K(x)^2dx\sum_{j=1}^{\infty}R(j) = o(1).
\end{aligned} \tag{A.15}
$$

We can show that $\mathbb{E}\,\hat{\Gamma}(\tau) - \Gamma(\tau) = o(1)$ by using very similar steps to the derivation of the asymptotic variance in the previous theorem. Again, the quantity can be split into four parts, where we use Lemmas A.2 and A.1 as well as A.5. This establishes (A.13). Thus, we proved that

$$(nh)\left[Var^*\left(\hat{m}^*(\tau) - \mathbb{E}^*(\hat{m}^*(\tau))\right)\right] \longrightarrow_p \sigma(\tau)^2 \left[R(0) + 2\sum_{m=1}^{\infty} R(m)\right] \int_{\mathbb{R}} K^2(u)du = \Gamma_0(\tau) + \Gamma(\tau),$$

which is the asymptotic variance expression $\sigma_{as}^2(\tau)$ as given in the theorem.

As in the proof of Theorem 4.1, we now show asymptotic normality of the bootstrap process using a similar blocking technique.

To create a situation of asymptotic independence of the blocks, we can use the following truncation. By construction, we get the following MA(∞) representation of $\{\xi_t^*\}_{t=1}^n$:

$$\xi_t^* = \gamma\xi_{t-1}^* + \nu_t^* = \sum_{j=0}^{\infty} \gamma^j \nu_{t-j}^*.$$

Assume that $M = M(n)$ is such that $M/\ell \to \infty$ as $n \to \infty$ and define $\xi_{t,M}^* = \sum_{j=0}^{M} \gamma^j \nu_{t-j}^*$. and, using this, the weighted average of interest becomes:

$$\bar{z}_{t,M}^*(\tau) \equiv \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau) z_{t,M}^* = \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau)\xi_{t,M}^*\hat{z}_t$$

Next, we show that the truncation error $\bar{W}_{n,M}^*(\tau) \equiv \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau)\hat{z}_t \sum_{j=M+1}^{\infty} \gamma^j \nu_{t-j}^*$ is negligible. Applying Markov's inequality twice, we have that

$$\begin{aligned}
\mathbb{E}(\mathbb{E}^*\left|\bar{W}_{n,M}^*(\tau)\right|) &\le \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau)(\mathbb{E}\,|\hat{z}_t|) \sum_{j=M+1}^{\infty} \gamma^j\, \mathbb{E}^*\left|\nu_{t-j}^*\right| \\
&\le (\mathbb{E}^*\,\nu_{t-j}^{*2})^{1/2} \max_{1\le t\le n} \mathbb{E}\,|\hat{z}_t| \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau) \sum_{j=M+1}^{\infty} \gamma^j \\
&\le \max_{1\le t\le n} \mathbb{E}\,|\hat{z}_t| \frac{1}{nh}\sum_{t=1}^{n} k_t(\tau)\frac{\gamma^M(1-\gamma^2)^{1/2}}{1-\gamma}.
\end{aligned}$$

As $\frac{1}{nh}\sum_{t=1}^{n} k_t(\tau) = O(1)$,

$$\begin{aligned}
\max_{1\le t\le n} \mathbb{E}\,|\hat{z}_t| &\le \max_{1\le t\le n} \mathbb{E}\,|\sigma_t u_t| + \left(\sup_{\tau} \mathbb{E}[\tilde{m}(t/n) - m(t/n)]^2\right)^{1/2} \\
&= O_p(1) + O_p\left(\max\left\{\tilde{h}^2, (n\tilde{h})^{-1/2}\right\}\right) = O_p(1),
\end{aligned} \tag{A.16}$$

and, as $M/\ell \to \infty$,

$$\frac{\gamma^M(1-\gamma^2)^{1/2}}{1-\gamma} = \gamma^M\left(\frac{1+\gamma}{1-\gamma}\right)^{1/2} = \theta^{M/\ell}\left(\frac{1+\theta^{1/\ell}}{1-\theta^{1/\ell}}\right)^{1/2} = \theta^{M/\ell} o(\ell^{1/2}) = o(1), \tag{A.17}$$

it follows that $\bar{W}_{n,M}^*(\tau) = o_p^*(1)$.

We split $(\hat{m}^*(\tau) - \mathbb{E}(\hat{m}^*(\tau)))$ into small and big blocks. As in the proof of Theorem 4.1, define the two sets $\mathcal{X}_i \equiv \{t : (i-1)(a+b)+1 \leq t \leq ia + (i-1)b\}$ and $\mathcal{Y}_i \equiv \{t : ia + (i-1)b + 1 \leq t \leq i(a+b)\}$ indexing the large and small blocks respectively. Define

$$\begin{aligned} X^*_{n,i}(\tau) &= \frac{1}{nh} \sum_{t \in \mathcal{X}_i} k_t(\tau) z^*_{t,M}, \\ Y^*_{n,i}(\tau) &= \frac{1}{nh} \sum_{t \in \mathcal{Y}_i} k_t(\tau) z^*_{t,M}, \end{aligned}$$

so that we obtain k blocks with $k = \left\lfloor \frac{n}{a+b} \right\rfloor$

$$\bar{z}^*_{t,M}(\tau) = \sum_{i=1}^{k} X^*_{n,i}(\tau) + \sum_{i=1}^{k} Y^*_{n,i}(\tau).$$

We again let $a(n) = o(n)$ and $b(n) = o(ah)$ while $b(n)/M \to \infty$ as $n \to \infty$.

We now show that the small blocks are negligible asymptotically:

$$\sqrt{nh} \sum_{i=1}^{k} Y^*_{n,i}(\tau) = o_p(1).$$

It is easy to see that $\mathbb{E}^* \left(\sqrt{nh} \sum_{i=1}^{k} Y^*_{n,i}(\tau) \right) = 0$. Now, we look at the variance. In comparison to above, the truncated version of the covariance has an additional term, but it can be bounded above by the original expression

$$\mathbb{E}^* \left(\xi^*_{t,M} \xi^*_{s,M} \right) = \theta^{\frac{k-l}{\ell}} \left(1 - \theta^{\frac{2(M+1)}{\ell}} \right) \leq \theta^{\frac{k-l}{\ell}}.$$

Replacing $\hat{z}_t$ in the same way as above yields the following four expressions:

$$\begin{aligned} Var^* \left(\sqrt{nh} \sum_{i=1}^{k} Y^*_{n,i}(\tau) \right) &= (nh)^{-1} \sum_{i=1}^{k} \sum_{t,s \in \mathcal{Y}_i} k_t(\tau) k_s(\tau) \hat{z}_t \hat{z}_s \, \mathbb{E}^* \left(\xi^*_{t,M} \xi^*_{s,M} \right) \\ &\leq (nh)^{-1} \sum_{i=1}^{k} \sum_{t,s \in \mathcal{Y}_i} k_t(\tau) k_s(\tau) \sigma_t u_t \sigma_s u_s(\tau) \theta^{\frac{k-l}{\ell}} \\ &\quad + (nh)^{-1} \sum_{i=1}^{k} \sum_{t,s \in \mathcal{Y}_i} k_t(\tau) k_s(\tau) \left(m(t/n) - \tilde{m}(t/n) \right) \left(m(s/n) - \tilde{m}(s/n) \right) \theta^{\frac{k-l}{\ell}} \\ &\quad + (nh)^{-1} \sum_{i=1}^{k} \sum_{t,s \in \mathcal{Y}_i} k_t(\tau) k_s(\tau) \sigma_t u_t \left(m(s/n) - \tilde{m}(s/n) \right) \theta^{\frac{k-l}{\ell}} \\ &\quad + (nh)^{-1} \sum_{i=1}^{k} \sum_{t,s \in \mathcal{Y}_i} k_t(\tau) k_s(\tau) \left(m(t/n) - \tilde{m}(t/n) \right) \sigma_s u_s \theta^{\frac{k-l}{\ell}} \\ &= Y^*_{1,n}(\tau) + Y^*_{2,n}(\tau) + Y^*_{3,n}(\tau) + Y^*_{4,n}(\tau) \end{aligned}$$

We now show all terms are asymptotically negligible. First note that

$$
\begin{aligned}
Y_{1,n}^{*} &= (nh)^{-1}\sum_{i=1}^{k}\sum_{t,s\in\mathcal{Y}_i} k_t(\tau)k_s(\tau)\sigma_t\sigma_s(\tau)u_tu_s\theta^{\frac{k-l}{\ell}} \\
&= (nh)^{-1}\sum_{i=1}^{k}\sum_{t,s\in\mathcal{Y}_i} k_t(\tau)k_s(\tau)\sigma_t\sigma_s(\tau)\,\mathbb{E}(u_tu_s)\theta^{\frac{k-l}{\ell}} \\
&\quad + (nh)^{-1}\sum_{i=1}^{k}\sum_{t,s\in\mathcal{Y}_i} k_t(\tau)k_s(\tau)\sigma_t\sigma_s(\tau)\left[u_tu_s-\mathbb{E}(u_tu_s)\right]\theta^{\frac{k-l}{\ell}} \\
&= Y_{1,n}^{*,I}(\tau)+Y_{1,n}^{*,II}(\tau)
\end{aligned}
$$

When we bound $\theta^{\frac{k-l}{\ell}}$ by 1, the $Y_{1,n}^{*,I}(\tau)$ coincides with the variance of the small blocks in the non-bootstrap case and can therefore be shown to converge to zero as $n\to\infty$ in exactly the same way. For $Y_{1,n}^{*,II}(\tau)$, we can rewrite the sums and use Lemma A.7(ii) for the two inner sums. Let $t_i = ia+(i-1)b$ as in the lemma. Then we obtain:

$$
\begin{aligned}
Y_{1,n}^{*,II}(\tau) =&(nh)^{-1}\sum_{i=1}^{k}\sum_{t,s\in\mathcal{Y}_i} k_t(\tau)k_s(\tau)\sigma_t\sigma_s(\tau)\left[u_tu_s-\mathbb{E}(u_tu_s)\right] \\
=&(nh)^{-1}\sum_{j=-b}^{b}\sum_{i=1}^{k}\sum_{t=1}^{b-|j|} k_{t_i+t}(\tau)k_{t_i+t+j}(\tau)\sigma_{t_i+t}\sigma_{t_i+t+j}\left[u_{t_i+t}u_{t_i+t+j}-\mathbb{E}(u_{t_i+t}u_{t_i+t+j})\right] \\
\le&\phi_n\sum_{j=-b}^{b}\beta_j+(nh)^{-1/2}\eta_n.
\end{aligned}
$$

Since $\phi_n\to 0$ as $n\to\infty$ and $\eta_n<\infty$ as well as $\sum_{j=-b}^{b}\beta_j<\infty$, we have that both terms in the last line go to zero and therefore, $Y_{1,n}^{*,II}(\tau)\to 0$ as $n\to\infty$.

Using Lemma A.4, expressions $Y_{2,n}^{*}(\tau)$, $Y_{3,n}^{*}(\tau)$ and $Y_{4,n}^{*}(\tau)$ can now be handled as $B_{2,n}^{*}(\tau)$, $B_{3,n}^{*}(\tau)$ and $B_{4,n}^{*}(\tau)$, respectively.

Finally, we show that large blocks follow a CLT and that the asymptotic variance coincides with the one given in the theorem. Using the Lindeberg CLT (see, e.g., Davidson (2002), Theorem 23.6), we can show that

$$
\sqrt{nh}\sum_{i=1}^{k}X_{n,i}^{*}(\tau)\xrightarrow{d^*}_p\mathcal{N}(0,\sigma_{as}^2(\tau)). \tag{A.18}
$$

Using the definition of $\mathcal{X}_i$ as above, we have that $\mathbb{E}^*\left(\sqrt{nh}\sum_{i=1}^{k}X_{n,i}^{*}(\tau)\right)=0$. Again, let n be such that $b(n) > M$.

$$
\begin{aligned}
Var^*\left(\sqrt{nh}\sum_{i=1}^{k} X^*_{n,i}(\tau)\right) &= (nh)^{-1}\sum_{i=1}^{k}\mathbb{E}^*(X^*_{n,i}(\tau))^2 \\
&= (nh)^{-1}\sum_{i=1}^{k}\sum_{t\in\mathcal{X}_i}\sum_{s\in\mathcal{X}_i} k_t(\tau)k_s(\tau)\,\mathbb{E}\left(z^*_{t,M}z^*_{s,M}\right) \\
&= (nh)^{-1}\sum_{i=1}^{k}\sum_{t\in\mathcal{X}_i} Var^*\left(k_t(\tau)\xi^*_{t,M}\hat{z}_t\right) \\
&\quad + 2(nh)^{-1}\sum_{i=1}^{k}\sum_{k,l\in\mathcal{X}_i;k<l} Cov^*\left(k_k(\tau)\xi^*_{k,M}\hat{z}_k, k_l(\tau)\xi^*_{l,M}\hat{z}_l\right) \\
&= D^*_n(\tau) + E^*_n(\tau)
\end{aligned}
$$

As above, we can show that $\Gamma_0(\tau)$ is the limit of $D^*_n(\tau)$ and $\Gamma(\tau)$ the limit of $E^*_n(\tau)$. We obtain for $D^*_n(\tau)$:

$$
\begin{aligned}
D^*_n(\tau) &= (nh)^{-1}\sum_{i=1}^{k}\sum_{t\in\mathcal{X}_i} k_t^2(\tau)\hat{z}_t Var^*\left(\xi^*_{t,M}\right) \\
&\le (nh)^{-1}\sum_{i=1}^{k}\sum_{t\in\mathcal{X}_i} k_t^2(\tau)\sigma_t^2 u_t^2 + (nh)^{-1}\sum_{i=1}^{k}\sum_{t\in\mathcal{X}_i} k_t^2(\tau)\left(m(t/n)-\tilde{m}(t/n)\right)^2 \\
&\quad + (nh)^{-1}\sum_{i=1}^{k}\sum_{t\in\mathcal{X}_i} k_t^2(\tau)\sigma_t u_t\left(m(t/n)-\tilde{m}(t/n)\right) \\
&= D^*_{1,n}(\tau) + D^*_{2,n}(\tau) + D^*_{3,n}(\tau),
\end{aligned}
$$

We observe that $D^*_{2,n}(\tau) = o_p(1)$ as well as $D^*_{3,n}(\tau) = o_p(1)$ following analogous arguments as for $A^*_{2,n}(\tau)$ and $A^*_{3,n}(\tau)$. $D^*_{1,n}(\tau)$ can be treated as $A^*_{1,n}(\tau)$ above, so the limit will indeed be $\Gamma_0(\tau)$. For $E^*_n(\tau)$, we proceed in analogy to part $B^*_n(\tau)$ and get the following four terms:

$$
\begin{aligned}
E^*_n(\tau) &\le 2(nh)^{-1}\sum_{i=1}^{k}\sum_{k,l\in\mathcal{X}_i;k<l} k_k(\tau)k_l(\tau)\sigma_k\sigma_l u_k u_l \theta^{\frac{l-k}{\ell}} \\
&\quad + 2(nh)^{-1}\sum_{i=1}^{k}\sum_{k,l\in\mathcal{X}_i;k<l} k_k(\tau)k_l(\tau)\left[m(k/n)-\tilde{m}(k/n)\right]\sigma_l u_l \theta^{\frac{l-k}{\ell}} \\
&\quad + 2(nh)^{-1}\sum_{i=1}^{k}\sum_{k,l\in\mathcal{X}_i;k<l} k_k(\tau)k_l(\tau)\left[m(l/n)-\tilde{m}(l/n)\right]\sigma_k u_k \theta^{\frac{l-k}{\ell}} \\
&\quad + 2(nh)^{-1}\sum_{i=1}^{k}\sum_{k,l\in\mathcal{X}_i;k<l} k_k(\tau)k_l(\tau)\left[m(k/n)-\tilde{m}(k/n)\right]\left[m(l/n)-\tilde{m}(l/n)\right]\theta^{\frac{l-k}{\ell}} \\
&= E^*_{1,n}(\tau) + E^*_{2,n}(\tau) + E^*_{3,n}(\tau) + E^*_{4,n}(\tau),
\end{aligned}
$$

The four parts can be treated as the corresponding parts of $B_n^*(\tau)$ such that overall, we have that

$$Var^* \left(\sqrt{nh} \sum_{i=1}^k X_{n,i}^*(\tau) \right) \longrightarrow_p \sigma(\tau)^2 \left[R(0) + 2 \sum_{m=1}^\infty R(m) \right] \int_{\mathbb{R}} K^2(u)du = \Gamma_0(\tau) + \Gamma(\tau),$$

The final step is to verify the Lindeberg condition. The first steps are analogous. For $\kappa > 0$,

$$\sum_{i=1}^k \mathbb{E}^* \left[\frac{X_{n,i}^*(\tau)^2}{\omega_n^{*2}} \mathbb{1}_{\left\{|X_{n,i}^*(\tau)/\omega_n| > \kappa\right\}} \right] = o(1) \tag{A.19}$$

with $\omega_n^{*2} = Var\left(\sum_{i=1}^k X_{n,i}^*\right) = O(nh)$ has to hold. We omit the detailed steps, since they are the same as in the proof of Theorem 4.1.

$$\sum_{i=1}^k \mathbb{E}^* \left[\frac{X_{n,i}^*(\tau)^2}{\omega_n^{*2}} \mathbb{1}_{\left\{|X_{n,i}^*(\tau)/\omega_n^*| > \kappa\right\}} \right] = \kappa^{-2} \sum_{i=1}^k \omega_n^{*-4} \, \mathbb{E}^* \left| X_{n,i}^*(\tau) \right|^4$$

For the fourth moment of big blocks we obtain using Minkowski's inequality

$$\begin{aligned}
\mathbb{E}^* |X_{n,i}^*(\tau)|^4 = \mathbb{E}^* \left| (nh)^{-1} \sum_{t \in \mathcal{X}_i} k_t(\tau) z_{t,M}^* \right|^4 &\leq (nh)^{-4} \left(\sum_{t \in \mathcal{X}_i} k_t(\tau) \left(\mathbb{E}^* \left| z_{t,M}^* \right|^4 \right)^{1/4} \right)^4 \\
&= \mathbb{E}^* \, \xi_{t,M}^4 \left((nh)^{-1} \sum_{t \in \mathcal{X}_i} k_t(\tau) \left| \hat{z}_t \right| \right)^4
\end{aligned}$$

As $\mathbb{E}^* \, \xi_{t,M}^4$ is bounded, and

$$(nh)^{-1} \sum_{t \in \mathcal{X}_i} k_t(\tau) \, \mathbb{E} \left| \hat{z}_t \right| \leq \max_{1 \leq t \leq n} \mathbb{E} \left| \hat{z}_t \right| (nh)^{-1} \sum_{t \in \mathcal{X}_i} k_t(\tau) = O_p(1) O\left(\frac{a}{nh} \right)$$

by (A.16), it follows that (A.19) holds by combining the above results. □

Proof of Theorem 4.3. In this proof we want to go from pointwise convergence which was shown in Theorems 4.1 and 4.2 to uniform convergence within $h(n)$-neighborhoods. We need to determine the covariance between estimates at two different points in a $h(n)$-neighborhood and show that it is correctly mimicked in the bootstrap world. Subsequently, we show stochastic equicontinuity and use the Cramér-Wold device to obtain the desired uniform convergence based on the foregoing theorems.

To derive the covariance expression given in the proof, we proceed in an analogy to the derivation of the asymptotic variance. Consider

$$\begin{aligned}
&Cov\left(N_{\tau_0,n}(\tau), N_{\tau_0,n}(\vartheta)\right) - \sigma(\tau)^2 \sum_{m=-\infty}^{\infty} R(m) \int_{\mathbb{R}} K(\omega)K(\omega+\vartheta-\tau)d\omega \\
&= (nh)Cov\left((nh)^{-1}\sum_{t=1}^{n} k_t(\tau_0+\tau h)y_t, (nh)^{-1}\sum_{t=0}^{n} k_t(\tau_0+\vartheta h)y_t\right) \\
&\quad - \sigma(\tau)^2 \sum_{m=-\infty}^{\infty} R(m) \int_{\mathbb{R}} K(\omega)K(\omega+\vartheta-\tau)d\omega \\
&= (nh)^{-1} \sum_{k=-n+1}^{n-1} R(k) \sum_{t=1}^{n-|k|} k_t(\tau_0+\tau h)k_{t+k}(\tau_0+\vartheta h)\sigma_t\sigma_{t+k} \\
&\quad - \sigma(\tau)^2 \sum_{m=-\infty}^{\infty} R(m) \int_{\mathbb{R}} K(\omega)K(\omega+\vartheta-\tau)d\omega \\
&= (nh)^{-2} \sum_{k=-n+1}^{n-1} R(k) \sum_{t=1}^{n-|k|} k_t(\tau_0+\tau h)k_{t+k}(\tau_0+\vartheta h)\left[\sigma_t\sigma_{t+k} - \sigma(\tau)^2\right] \\
&\quad + (nh)^{-1}\sigma(\tau)^2 \sum_{k=-n+1}^{n-1} R(k) \sum_{t=1}^{n-|k|} \left[k_t(\tau_0+\tau h)k_{t+k}(\tau_0+\vartheta h) - k_t(\tau_0+\tau h)k_t(\tau_0+\vartheta h)\right] \\
&\quad + (nh)^{-1}\sigma(\tau)^2 \sum_{k=-n+1}^{n-1} R(k) \left[\sum_{t=1}^{n-|k|} k_t(\tau_0+\tau h)k_t(\tau_0+\vartheta h) - \int_{\mathbb{R}} K(\omega)K(\omega+\vartheta-\tau)d\omega\right] \\
&\quad + \sigma(\tau)^2 \left[\sum_{k=-n+1}^{n-1} R(k) - \sum_{m=-\infty}^{\infty} R(m)\right] \int_{\mathbb{R}} K(\omega)K(\omega+\vartheta-\tau)d\omega \\
&= H_{1,n} + H_{2,n} + H_{3,n} + H_{4,n}
\end{aligned}$$

All parts can be shown to be asymptotically negligible. Let us start with $H_{1,n}$. By Lemmas A.1 and A.2, we have that

$$\begin{aligned}
|H_{1,n}| &\leq (nh)^{-1} \sum_{k=-n+1}^{n-1} |R(k)| \sum_{t=1}^{n-|k|} k_t(\tau_0+\tau h)k_{t+k}(\tau_0+\vartheta h)\left|\sigma_t\sigma_{t+k} - \sigma(\tau)^2\right| \\
&\leq (nh)^{-1} \sum_{k=-n+1}^{n-1} |R(k)|\, C(nh)h \leq h \sum_{k=-n+1}^{n-1} R(k) \leq Ch.
\end{aligned}$$

Hence, $H_{1,n} \to 0$ as $n \to \infty$. Similarly, for $H_{2,n}$ we can apply Lemma A.1 and Lemma A.5 together with the Lipschitz property of the Kernel function

$$\begin{aligned}
|H_{2,n}| &\le (nh)^{-1}\sigma(\tau)^2 \sum_{k=-n+1}^{n-1} |R(k)| \sum_{t=1}^{n-|k|} |k_t(\tau_0+\tau h)k_{t+k}(\tau_0+\vartheta h) - k_t(\tau_0+\tau h)k_t(\tau_0+\vartheta h)| \\
&\le (nh)^{-1}\sigma(\tau)^2 \sum_{k=-n+1}^{n-1} |R(k)| \sum_{t=1}^{n-|k|} k_t(\tau_0+\tau h)\,|k_{t+k}(\tau_0+\vartheta h) - k_t(\tau_0+\vartheta h)| \\
&\le (nh)^{-2}\sigma(\tau)^2 \sum_{k=-n+1}^{n-1} k\,|R(k)| \sum_{t=1}^{n-|k|} k_t(\tau_0+\tau h) \\
&\le (nh)^{-1}\sigma(\tau)^2 \sup_s K(s) \sum_{k=-n+1}^{n-1} k\,|R(k)| \le C(nh)^{-1}
\end{aligned}$$

such that we get that $H_{2,n} \to 0$ as $n \to \infty$.

Next we look at $H_{3,n}$. We can write

$$\begin{aligned}
|H_{3,n}| \le{}& 2(nh)^{-1}\sigma(\tau)^2 \sum_{k=-n+1}^{n-1} |R(k)| \left| \sum_{t=|k|}^{n-|k|} k_t(\tau_0+\tau h)k_t(\tau_0+\vartheta h) - \sum_{t=1}^{n} k_t(\tau_0+\tau h)k_t(\tau_0+\vartheta h) \right| \\
&+ 2\sigma(\tau)^2 \sum_{k=-n+1}^{n-1} |R(k)| \left| (nh)^{-1} \sum_{t=1}^{n} k_t(\tau_0+\tau h)k_t(\tau_0+\vartheta h) - \int_{\mathbb{R}} K(\omega)K(\omega+\vartheta-\tau)d\omega \right| \\
={}& H_{3,n}^{I} + H_{3,n}^{II}.
\end{aligned}$$

For $H_{3,n}^{I}$ we have

$$\begin{aligned}
\left|H_{3,n}^{I}\right| &= 2(nh)^{-1}\sigma(\tau)^2 \sum_{k=-n+1}^{n-1} |R(k)| \left| \sum_{t=n-|k|+1}^{n} k_t(\tau_0+\tau h)k_t(\tau_0+\vartheta h) \right| \\
&\le 2(nh)^{-1} \sup_s \left\{K(s)^2\right\} \sigma(\tau)^2 \sum_{k=-n+1}^{n-1} |k|\,|R(k)| \le C(nh)^{-1}.
\end{aligned}$$

The first inequality follows from the fact that the sum over the Kernel functions has $|k|$ elements. An application of Lemma A.5 yields the desired result.
Furthermore $\left|H_{3,n}^{II}\right| \le C\left|(nh)^{-1}k_t(\tau_0+\tau h)k_t(\tau_0+\vartheta h) - \int_{\mathbb{R}} K(\omega)K(\omega+\vartheta-\tau)d\omega\right| = o(1)$ by Lemma A.3, hence $H_{3,n} \to 0$ as $n \to \infty$.

Finally, to deal with $H_{4,n}$, define m_n such that $m_n \to \infty$ and $m_n/\ell \to 0$ as $n \to \infty$. Then we have that

$$\begin{aligned}
|H_{4,n}| &\le \sigma(\tau)^2 \left| \sum_{k=-n+1}^{n-1} R(k) - \sum_{m=-\infty}^{\infty} R(m) \right| \int_{\mathbb{R}} K(\omega)K(\omega+\vartheta-\tau)d\omega \\
&\le \sigma(\tau)^2 2 \sum_{k=n}^{\infty} |R(k)| \int_{\mathbb{R}} K(\omega)K(\omega+\vartheta-\tau)d\omega = o(1)
\end{aligned}$$

The last part follows from the fact that the tail sum of the autocovariances goes to zero as $n \to \infty$. Hence, we have verified the asymptotic covariance within h-neighborhoods given in the theorem.

The next step is to show that it is the same for the bootstrap case.

$$
\begin{aligned}
&Cov^* \left(N^*_{\tau_0,n}(\tau), N^*_{\tau_0,n}(\vartheta) \right) \\
&= (nh)^{-1} \sum_{t=1}^{n} \sum_{k=-n+1}^{n-1} k_t(\tau_0 + \tau h) k_{t+k}(\tau_0 + \tau h) \hat{z}_t \hat{z}_{t+k} \theta^{\frac{k}{\ell}} \\
&= (nh)^{-1} \sum_{k=-n+1}^{n-1} \theta^{\frac{k}{\ell}} \sum_{t=1}^{n} k_t(\tau_0 + \tau h) k_{t+k}(\tau_0 + \tau h) \sigma_t \sigma_{t+k} u_t u_{t+k} \\
&\quad + (nh)^{-1} \sum_{k=-n+1}^{n-1} \theta^{\frac{k}{\ell}} \sum_{t=1}^{n} k_t(\tau_0 + \tau h) k_{t+k}(\tau_0 + \tau h) \left(m(t/n) - \tilde{m}(t/n) \right) \left(m((t+k)/n) - \tilde{m}((t+k)/n) \right) \\
&\quad + (nh)^{-1} \sum_{k=-n+1}^{n-1} \theta^{\frac{k}{\ell}} \sum_{t=1}^{n} k_t(\tau_0 + \tau h) k_{t+k}(\tau_0 + \tau h) \left(m(t/n) - \tilde{m}(t/n) \right) \sigma_{t+k} u_{t+k} \\
&\quad + (nh)^{-1} \sum_{k=-n+1}^{n-1} \theta^{\frac{k}{\ell}} \sum_{t=1}^{n} k_t(\tau_0 + \tau h) k_{t+k}(\tau_0 + \tau h) \left(m((t+k)/n) - \tilde{m}((t+k)/n) \right) \sigma_t u_t \\
&= I^*_{1,n} + I^*_{2,n} + I^*_{3,n} + I^*_{4,n}
\end{aligned}
$$

$I^*_{1,n}$ is the dominating part and can, in analogy to $B^*_{n,1}$, be shown to have the desired limit with the help of Lemma A.7(i). Then, it remains to show that $I^*_{2,n}$, $I^*_{3,n}$ and $I^*_{4,n}$ are negligible asymptotically. But this follows using the same steps as for $B^*_{2,n}$, $B^*_{3,n}$ and $B^*_{4,n}$, respectively.

For the remainder of the proof, it is more convenient to work with centered versions of $N_{\tau_0,n}$ and $N^*_{\tau_0,n}$, respectively, as in the proofs of the previous two theorems. We first show that these centered quantities are equivalent to bias corrected versions of $N_{\tau_0,n}$ and $N^*_{\tau_0,n}$. More specifically, define

$$
\begin{aligned}
W_{\tau_0,n}(\tau) &\equiv N_{\tau_0,n}(\tau) - \mathbb{E}\left(N_{\tau_0,n}(\tau) \right), \\
W^*_{\tau_0,n}(\tau) &\equiv N^*_{\tau_0,n}(\tau) - \mathbb{E}^*\left(N^*_{\tau_0,n}(\tau) \right).
\end{aligned}
$$

The next step is to show that

$$
\begin{aligned}
&\sup_{\tau \in [-1,1]} \left| \mathbb{E}\left(N_{\tau_0,n}(\tau) \right) - B_{as}(\tau_0) \right| = o(1), \\
&\sup_{\tau \in [-1,1]} \left| \mathbb{E}^*\left(N^*_{\tau_0,n}(\tau) \right) - B_{as}(\tau_0) \right| = o_p(1).
\end{aligned}
$$

The statements follow from Lemma A.8(i) and Lemma A.8(ii), respectively, and they show that we can work with the centered processes $\{W_{\tau_0,n}(\tau)\}_{\tau\in[-1,1]}$ and $\left\{W^*_{\tau_0,n}(\tau)\right\}_{\tau\in[-1,1]}$.

For these processes we have univariate convergence to the same limiting normal distribution as in Theorems 4.1 and 4.2 which is

$$
\begin{aligned}
(nh)^{1/2} W_{\tau_0,n}(\tau) &\xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{as}\right) \\
(nh)^{1/2} W^*_{\tau_0,n}(\tau) &\xrightarrow{d^*} \mathcal{N}\left(0, \sigma^2_{as}\right).
\end{aligned}
$$

Applying the Cramér-Wold device to go from univariate to multivariate convergence, we obtain the limiting distribution of the vector $(W_{\tau_0,n}(\tau_1),...,W_{\tau_0,n}(\tau_m))'$ and $(W^*_{\tau_0,n}(\tau_1),...,W^*_{\tau_0,n}(\tau_m))'$, $\tau_1,...,\tau_m \in [-1,1]$, $m \in \mathbb{N}$. It remains to show stochastic equicontinuity of $W_{\tau_0,n}(\tau)$ and $W^*_{\tau_0,n}(\tau)$, which is given by the following statements. For $\kappa > 0$, $\eta > 0$, there exists $\lambda > 0$ and $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}\left(\sup_{|\tau-\vartheta|\le\lambda} |W_{\tau_0,n}(\tau) - W_{\tau_0,n}(\vartheta)| > \kappa\right) < \eta \qquad \forall\, n \ge n_0$$

$$\mathbb{P}\left[\mathbb{P}^*\left(\sup_{|\tau-\vartheta|\le\lambda} \left|W^*_{\tau_0,n}(\tau) - W^*_{\tau_0,n}(\vartheta)\right| > \kappa\right) < \eta\right] > 1-\eta \qquad \forall\, n \ge n_0.$$

We apply Theorem 12.3 in Billingsley (1968) to obtain tightness of $W_{\tau_0,n}(\cdot)$ and $W^*_{\tau_0,n}(\cdot)$ and with tightness, we have stochastic equicontinuity, which is sufficient to show uniform convergence as stated in the theorem. We first argue that the two conditions of Theorem 12.3 in Billingsley (1968) are satisfied. The condition of tightness of $W_{\tau_0,n}(0)$ is already established. To see why, recall the definition of $W_{\tau_0,n}(\cdot)$ as a centered version of the neighborhood $N_{\tau_0,n}(\cdot)$. At the point $\tau = 0$ we only look at the centered estimator at a specific point τ_0 and since pointwise weak convergence has been established, we immediately obtain tightness of $N_{\tau_0,n}(0)$ and, in turn, $W_{\tau_0,n}(0)$. The same holds for tightness of $W^*_{\tau_0,n}(0)$ Next, we are going to verify the moment condition for $W_{\tau_0,n}(\cdot)$. To do so, consider

$$W_{\tau_0,n}(\tau) - W_{\tau_0,n}(\vartheta) = (nh)^{-1/2}\sum_{t=1}^{n}\left(k_t(\tau_0+\tau h) - k_t(\tau_0+\vartheta h)\right) z_t$$

such that for the expected squared fluctuations we have

$$\begin{aligned}&\mathbb{E}\,|W_{\tau_0,n}(\tau) - W_{\tau_0,n}(\vartheta)|^2 \\ &\quad\le (nh)^{-1}\sum_{t=1}^{n}\sum_{k=-n+1}^{n-1} |k_t(\tau_0+\tau h) - k_t(\tau_0+\vartheta h)|\,|k_{t-k}(\tau_0+\tau h) - k_{t-k}(\tau_0+\vartheta h)|\,\sigma_t\sigma_{t-k}R(k)\end{aligned}$$

Applying Lemma A.1 and A.2 as well as using the Lipschitz property of the Kernel function gives us the moment condition

$$\mathbb{E}\,|W_{\tau_0,n}(\tau) - W_{\tau_0,n}(\vartheta)|^2 \le C\,|\tau-\vartheta|^2. \tag{A.20}$$

To verify this condition for $W^*_{\tau_0,n}(\cdot)$, consider

$$W^*_{\tau_0,n}(\tau) - W^*_{\tau_0,n}(\vartheta) = (nh)^{-1/2}\sum_{t=1}^{n}\left(k_t(\tau_0+\tau h) - k_t(\tau_0+\vartheta h)\right) z^*_t.$$

To show that

$$\mathbb{E}^*\left|W^*_{\tau_0,n}(\tau) - W^*_{\tau_0,n}(\vartheta)\right|^2 \le C\,|\tau-\vartheta|^2, \tag{A.21}$$

we define the index set $\mathcal{I}_{n,h}(x) \equiv [n(x-h), n(x+h)]$ for which the function $k_t(x)$ is non-zero such that we can reduce the following sums to sums over non-zero elements only.

Let $\mathcal{I} = \mathcal{I}_{n,h}(\tau_0 + \tau h) \cup \mathcal{I}_{n,h}(\tau_0 + \vartheta h)$, then

$$
\begin{aligned}
&\mathbb{E}^* \left| W^*_{\tau_0,n}(\tau) - W^*_{\tau_0,n}(\vartheta) \right|^2 \\
&\quad \le (nh)^{-1} \sum_{t=1}^{n} \sum_{k=-n+1}^{n-1} \left| k_t(\tau_0 + \tau h) - k_t(\tau_0 + \vartheta h) \right| \left| k_{t-k}(\tau_0 + \tau h) - k_{t-k}(\tau_0 + \vartheta h) \right| \hat{z}_t \hat{z}_{t-k} \theta^{\frac{k}{\ell}} \\
&\quad \le (nh)^{-1} \sum_{t\in\mathcal{I}} \sum_{k=-n+1}^{n-1} \left| k_t(\tau_0 + \tau h) - k_t(\tau_0 + \vartheta h) \right| \left| k_{t-k}(\tau_0 + \tau h) - k_{t-k}(\tau_0 + \vartheta h) \right| \hat{z}_t \hat{z}_{t-k} \theta^{\frac{k}{\ell}} \\
&\quad \le \left| \tau - \vartheta \right|^2 (nh)^{-1} \sum_{k=-n+1}^{n-1} \theta^{\frac{k}{\ell}} \sum_{t\in\mathcal{I}} \hat{z}_t \hat{z}_{t-k} \\
&\quad \le \left| \tau - \vartheta \right|^2 (nh)^{-1} \sum_{k=-n+1}^{n-1} \theta^{\frac{k}{\ell}} \sum_{t\in\mathcal{I}} \sigma_t \sigma_{t-k} u_t u_{t-k} \\
&\qquad + \left| \tau - \vartheta \right|^2 (nh)^{-1} \sum_{k=-n+1}^{n-1} \theta^{\frac{k}{\ell}} \sum_{t\in\mathcal{I}} \sigma_t u_t \left(m((t-k)/n) - \tilde{m}((t-k)/n) \right) \\
&\qquad + \left| \tau - \vartheta \right|^2 (nh)^{-1} \sum_{k=-n+1}^{n-1} \theta^{\frac{k}{\ell}} \sum_{t\in\mathcal{I}} \sigma_{t-k} u_{t-k} \left(m(t/n) - \tilde{m}(t/n) \right) \\
&\qquad + \left| \tau - \vartheta \right|^2 (nh)^{-1} \sum_{k=-n+1}^{n-1} \theta^{\frac{k}{\ell}} \sum_{t\in\mathcal{I}} \left(m(t/n) - \tilde{m}(t/n) \right) \left(m((t-k)/n) - \tilde{m}((t-k)/n) \right) \\
&\quad = J^*_{1,n} + J^*_{2,n} + J^*_{3,n} + J^*_{4,n}
\end{aligned}
$$

We observe that there is a tight link between $J^*_{2,n}$ and $B^*_{2,n}$, because the sum over the set $\mathcal{I}$ makes sure that we sum over at most (nh) elements. This was ensured in $B^*_{2,n}$ by the kernel functions. So, $J^*_{2,n}$ can be shown to be asymptotically negligible following the same steps as for $B^*_{2,n}$. The same correspondence is true for $J^*_{3,n}$ and $B^*_{3,n}$ as well as $J^*_{4,n}$ and $B^*_{4,n}$. They can be handled analogously such that we are left with $J^*_{1,n}$. Here, we can make use of Lemma A.7(i) in a similar way as before by adding and subtracting $\mathbb{E}\, u_t u_{t-k}$ in the original expression

$$
\begin{aligned}
J^*_{1,n} = &\left| \tau - \vartheta \right|^2 (nh)^{-1} \sum_{k=-n+1}^{n-1} \theta^{\frac{k}{\ell}} \sum_{t\in\mathcal{I}} \sigma_t \sigma_{t-k} \left(u_t u_{t-k} - \mathbb{E}\, u_t u_{t-k} \right) \\
&+ \left| \tau - \vartheta \right|^2 (nh)^{-1} \sum_{k=-n+1}^{n-1} \theta^{\frac{k}{\ell}} \sum_{t\in\mathcal{I}} \sigma_t \sigma_{t-k}\, \mathbb{E}\, u_t u_{t-k},
\end{aligned}
$$

where the first part converges to zero as $n \to \infty$ and the sum in the second part is bounded. Thus, we have established (A.20) and (A.21) which is sufficient to show stochastic equicontinuity. This completes the proof. $\square$

B Appendix: Proofs of auxiliary lemmas

Proof of Lemma A.1. Note that the sum over $k_t k_{t+i}$ as well as the sum over $k_t^2 k_{t+i}^2$ has at most $2nh$ non-zero elements as $k_t = K\left(\frac{t/n-\tau}{h}\right) = 0$ whenever $|t/n - \tau| > h$. This holds for any $i \in \mathbb{Z}$, so it also holds for the maximum. Furthermore, given Assumption 4 we know that $\sup_s K(s) < \infty$

holds. $\square$

Proof of Lemma A.2. Given the bounds on $|l/n - \tau|$ and $|k/n - \tau|$, we can show the validity of the Lemma in the following way:

$$\left|\sigma_k \sigma_l - \sigma(\tau)^2\right| = \left|\sigma_k \sigma_l - \sigma(\tau)^2 - \sigma_k \sigma(\tau) + \sigma_k \sigma(\tau)\right| \leq C_3 \left(|\sigma_k| + |\sigma(\tau)|\right) h = O(h).$$

$\square$

Proof of Lemma A.3. The proofs for parts (i) and (ii) run analogously, so we just show the second case. We can take sequential limits and first evaluate $h^{-1} \lim_{n\to\infty} n^{-1} K\left(h^{-1}(t/n - \tau)\right)^2$ and then take the limit of $h \to 0$. In general, we have the following Riemann-sum approximation of an integral $\lim_{n\to\infty} \sum_{t=1}^{n} f(\frac{t}{n})\frac{1}{n} = \int_0^1 f(t)dt$. This leads us to

$$h^{-1} \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} K\left(\frac{t/n - \tau}{h}\right)^2 = h^{-1} \int_0^1 K\left(\frac{t-\tau}{h}\right)^2 dt.$$

Now substitute $\omega = h^{-1}(t - \tau)$. Then, we have $d\omega = \frac{dt}{h}$ and the lower integration bound becomes $h^{-1}(1 - \tau)$, while the upper bound will be $-\tau/h$.

$$h^{-1} \int_0^1 K\left(\frac{t-\tau}{h}\right)^2 dt = h^{-1} \int_{-\frac{\tau}{h}}^{\frac{1-\tau}{h}} K(\omega)^2 \, h d\omega,$$

letting $h \to 0$ gives

$$\lim_{h\to 0} h^{-1} \int_{-\frac{\tau}{h}}^{\frac{1-\tau}{h}} K(\omega)^2 \, h d\omega = \int_{\mathbb{R}} K(\omega)^2 \, d\omega,$$

which concludes parts (i) and (ii). For part (iii), we first have that

$$\frac{1}{h} \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} K\left(\frac{t/n - \tau_0 - \tau h}{h}\right) K\left(\frac{t/n - \tau_0 - \vartheta h}{h}\right) = \frac{1}{h} \int_0^1 K\left(\frac{t - \tau_0 - \tau h}{h}\right) K\left(\frac{t/n - \tau_0 - \vartheta h}{h}\right) dt.$$

First, let $\eta = t/n$ such that $dt = h d\eta$. The lower integration bound stays 0, the upper bound becomes h^{-1}. Then

$$\frac{1}{h} \int_0^1 K\left(\frac{t - \tau_0 - \tau h}{h}\right) K\left(\frac{t/n - \tau_0 - \vartheta h}{h}\right) dt = \int_0^{1/h} K\left(\eta - \frac{\tau_0 + \tau h}{h}\right) K\left(\eta - \frac{\tau_0 + \vartheta h}{h}\right) d\eta$$

Second, let $\omega = \eta - \frac{\tau_0 + \tau h}{h}$. Then, $d\omega = d\eta$ and the lower bound becomes $-\frac{\tau_0}{h} + \tau$, while the upper will be $\frac{(1-\tau_0)}{h} + \tau$. We have

$$\int_0^{1/h} K\left(\eta - \frac{\tau_0 + \tau h}{h}\right) K\left(\eta - \frac{\tau_0 + \vartheta h}{h}\right) d\eta = \int_{-\frac{\tau_0}{h}+\tau}^{\frac{(1-\tau_0)}{h}+\tau} K(\omega) \, K(\omega + \vartheta - \tau) d\omega$$

Finally, for $h \to 0$ the integration bounds go to $-\infty$ and ∞, respectively, and the resulting integral coincides with the one given in the Lemma. $\square$

Proof of Lemma A.4. Let $\delta > 0$. By Markov's inequality we have that

$$\begin{aligned}\mathbb{P}\left(\frac{1}{n}\sum_{t=1}^{n}\left(\tilde{m}(t/n)-m(t/n)\right)^2>\delta\right)&\leq\frac{\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left(\tilde{m}(t/n)-m(t/n)\right)^2}{\delta}\\&\leq\frac{\sup_t\mathbb{E}\left(\tilde{m}(t/n)-m(t/n)\right)^2}{\delta}\\&\leq\frac{\sup_{\tau\in[0,1]}\mathbb{E}\left(\tilde{m}(\tau)-m(\tau)\right)^2}{\delta}\end{aligned}$$

Therefore, it is sufficient to bound the supremum of variance and the squared bias of the estimator over all $\tau \in [0,1]$ to obtain the desired result.

$$\begin{aligned}\sup_\tau\mathbb{E}\left(\tilde{m}(\tau)-m(\tau)\right)^2&=\sup_\tau Var(\tilde{m}(\tau))+\sup_\tau\left[\mathbb{E}\left(\tilde{m}(\tau)\right)-m(\tau)\right]^2\\&=O\left((n\tilde{h})^{-1}\right)+O\left(\tilde{h}^4\right)=O\left(\max\left\{\tilde{h}^4,(n\tilde{h})^{-1}\right\}\right)\end{aligned}$$

The bound on the variance is taken from the derivation of the asymptotic variance in the proof of Theorem 4.1. The bound is uniform over τ, since the only term in the asymptotic variance which is dependent on τ is $\sigma(\tau)^2$ and its supremum taken over $\tau \in [0,1]$ is bounded. For the bias part, consider

$$\mathbb{E}\left(\tilde{m}(\tau)\right)-m(\tau)=(n\tilde{h})^{-1}\sum_{t=1}^{n}K\left(\frac{t/n-\tau}{\tilde{h}}\right)m(t/n)-m(\tau)$$

To handle this expression, we can use a bound on the approximation of an integral by a sum, given in Bühlmann (1998) and restated here (see the proof of Theorem 3.2, equation (6.5)). Let $g(\cdot)$ be a continuous and Riemann-integrable function, then

$$\left|n^{-1}\sum_{t=1}^{n}g(t/n)-\int_0^1 g(z)dz\right|\leq\sup_{|x-y|\leq n^{-1}}|g(x)-g(y)|. \tag{B.1}$$

In our case, consider $g(x)=K\left(\frac{x-\tau}{\tilde{h}}\right)m(x)$. With the help of the bound we obtain

$$\begin{aligned}&\left|n^{-1}\sum_{t=1}^{n}K\left(\frac{t/n-\tau}{\tilde{h}}\right)m(t/n)-\int_0^1 K(u)m(u\tilde{h}+\tau)du\right|\\&\leq\sup_{|x-y|\leq n^{-1}}\left|K\left(\frac{x-\tau}{\tilde{h}}\right)m(x)-K\left(\frac{y-\tau}{\tilde{h}}\right)m(y)\right|\\&\leq\sup_{|x-y|\leq n^{-1}}\left|K\left(\frac{x-\tau}{\tilde{h}}\right)\right||x-y|+|m(y)|\frac{|x-y|}{\tilde{h}}\leq C_1n^{-1}+C_2(n\tilde{h})^{-1}\end{aligned}$$

To see the correspondence of arguments, we have $u=(t/n-\tau)/\tilde{h}$ so that $t/n=u\tilde{h}+\tau$. We can now bound the supremum of the bias term using a second order Taylor approximation of $m(u\tilde{h}+\tau)$ with $\tau \in [0,1]$,

$$m(u\tilde{h}+\tau)\approx m(\tau)+m'(\tau)(-u\tilde{h})+m''(\tau)(-u\tilde{h})^2/2$$

and replace $m(u\tilde{h}+\tau)$ in the integral above by his approximation:

$$
\begin{aligned}
\sup_{\tau}\left|\mathbb{E}\left(\tilde{m}(\tau)\right)-m(\tau)\right| &= \sup_{\tau}\left|(n\tilde{h})^{-1}\sum_{t=1}^{n}K\left(\frac{t/n-\tau}{\tilde{h}}\right)m(t/n)-m(\tau)\right| \\
&\leq \sup_{\tau}\left|n^{-1}\sum_{t=1}^{n}K\left(\frac{t/n-\tau}{\tilde{h}}\right)m(t/n)-\int_0^1 K(u)m(u\tilde{h}+\tau)du\right| \\
&\quad+\sup_{\tau}\left|\int_0^1 K(u)m(u\tilde{h}+\tau)du-m(\tau)\right| \\
&\leq C_1 n^{-1}+C_2(n\tilde{h})^{-1}+\sup_{\tau}\left|m''(\tau)\tilde{h}^2\int_0^1 K(u)u^2du/2\right| \\
&= O\left(\max\left\{n^{-1},(n\tilde{h})^{-1},\tilde{h}^2\right\}\right)=O\left(\tilde{h}^2\right),
\end{aligned}
$$

where the first part of the third line comes from the integral approximation bound and the other part follows from the Taylor approximation as follows:

$$
\begin{aligned}
&\sup_{\tau}\left|\int_0^1 K(u)m(u\tilde{h}+\tau)du-m(\tau)\right| \\
&\approx \sup_{\tau}\left|\int_0^1 K(u)\left[m(\tau)+m'(\tau)(-u\tilde{h})+m''(\tau)(-u\tilde{h})^2/2\right]du-m(\tau)\right| \\
&= \sup_{\tau}\left|m(\tau)\int_0^1 K(u)du-\tilde{h}m'(\tau)\int_0^1 K(u)udu+\tilde{h}^2 m''(\tau)\int_0^1 K(u)u^2/2du-m(\tau)\right| \\
&= \sup_{\tau}\left|m''(\tau)\tilde{h}^2\int_0^1 K(u)u^2du/2\right|.
\end{aligned}
$$

The first integral over the Kernel function equals 1 by Assumption 4 and thus, this part cancels with $m(\tau)$. The integral in the second term equals $\mathbb{E}\left(K(\cdot)\right)=0$ by symmetry of the Kernel function and it remains the last integral. By Assumption 1 we know that $\sup_{\tau}|m''(\tau)|<\infty$ which shows that it is of order $O\left(\tilde{h}^2\right)$. This completes the proof. □

Proof of Lemma A.5.

$$
\sum_{k=1}^{\infty}k\left|R(k)\right|=\sum_{k=1}^{\infty}k\left|\sum_{l=0}^{\infty}\psi_l\psi_{l+k}\right|\leq\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}k\left|\psi_l\right|\left|\psi_k\right|=\left(\sum_{l=0}^{\infty}\left|\psi_l\right|\right)\left(\sum_{k=0}^{\infty}k\left|\psi_k\right|\right)<\infty,
$$

□

Proof of Lemma A.6. Using the Lipschitz continuity of the Kernel function $K(\cdot)$ we obtain

$$
\begin{aligned}
\left|k_t^2(\tau)-k_s^2(\tau)\right| &= \left|k_t^2(\tau)-k_s^2(\tau)+k_t(\tau)k_s(\tau)-k_t(\tau)k_s(\tau)\right| \\
&= \left|k_t(\tau)\left[k_s(\tau)-k_t(\tau)\right]-k_s(\tau)\left[k_s(\tau)-k_t(\tau)\right]\right|\leq\frac{C_1|t-s|}{nh}\left|k_s(\tau)-k_t(\tau)\right|\leq\frac{C_2|t-s|^2}{(nh)^2}.
\end{aligned}
$$

□

Proof of Lemma A.7. We only give a proof for statement (i) of the lemma as part (ii) follows with minor adjustments. As in Jansson (2002), we have that

$$
\begin{aligned}
u_t u_{t+i} &= \left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right)\left(\sum_{m=0}^{\infty} \psi_m \varepsilon_{t+i-m}\right) \\
&= \sum_{j=0}^{\infty} \psi_j \psi_{j+i} \varepsilon_{t-j}^2 + \sum_{j=0}^{\infty}\sum_{m=0}^{\infty} \mathbb{1}_{m \neq j+i} \psi_j \psi_m \varepsilon_{t-j} \varepsilon_{t+i-m},
\end{aligned}
$$

such that

$$
u_t u_{t+i} - \mathbb{E}\, u_t u_{t+i} = \sum_{j=0}^{\infty} \psi_j \psi_{j+i} (\varepsilon_{t-j}^2 - 1) + \sum_{j=0}^{\infty}\sum_{m=0}^{\infty} \mathbb{1}_{m \neq j+i} \psi_j \psi_m \varepsilon_{t-j} \varepsilon_{t+i-m},
$$

and

$$
\begin{aligned}
A_{n,h} &= (nh)^{-1} \sum_{t=1}^{n-i} k_t(\tau) k_{t+i}(\tau) \sigma_t \sigma_{t+i} \left(u_t u_{t+i} - \mathbb{E}\, u_t u_{t+i}\right) \\
&= \sum_{j=0}^{\infty} \psi_j \psi_{j+i} \left[(nh)^{-1} \sum_{t=1}^{n-i} k_t(\tau) k_{t+i}(\tau) \sigma_t \sigma_{t+i} (\varepsilon_{t-j}^2 - 1)\right] \\
&\quad + \sum_{j=0}^{\infty}\sum_{m=0}^{\infty} \mathbb{1}_{m \neq j+i} \psi_j \psi_m (nh)^{-1} \sum_{t=1}^{n-i} k_t(\tau) k_{t+i}(\tau) \sigma_t \sigma_{t+i} \varepsilon_{t-j} \varepsilon_{t+i-m}.
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathbb{E}\,|A_{n,h}| &\leq \sup_{j \geq 0} \max_{0 \leq i \leq n-1} \mathbb{E} \left|(nh)^{-1} \sum_{t=1}^{n-i} k_t(\tau) k_{t+i}(\tau) \sigma_t \sigma_{t+i} (\varepsilon_{t-j}^2 - 1)\right| \left(\sum_{j=0}^{\infty} |\psi_j|\, |\psi_{j+i}|\right) \\
&\quad + (nh)^{-1/2} \sup_{j,m \geq 0} \max_{0 \leq i \leq n-1} \mathbb{E} \left|(nh)^{-1/2} 1(m \neq j+i) \sum_{t=1}^{n-i} k_t(\tau) k_{t+i}(\tau) \sigma_t \sigma_{t+i} \varepsilon_{t-j} \varepsilon_{t+i-m}\right| \\
&\quad \times \sum_{j=0}^{\infty}\sum_{m=0}^{\infty} |\psi_j|\, |\psi_m| =: \phi_n \beta_i + (nh)^{-1/2} \eta_n.
\end{aligned}
$$

First note that

$$
\sum_{i=0}^{\infty} \beta_i \leq \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} |\psi_j|\, |\psi_{j+i}| \leq \left(\sum_{j=0}^{\infty} |\psi_j|\right)^2 < \infty.
$$

Next, let $w_t = \varepsilon_t^2 - 1$. By Assumption 3, $\mathbb{E}\,|w_t|^p \leq \mathbb{E}\,|\varepsilon_t|^{2p} < \infty$ for some $p > 1$. This implies that w_t is uniformly integrable(cf. Davidson, 2002, Theorem 12.10), from which we can conclude that for every $\epsilon > 0$, there exists a $\lambda_\epsilon > 0$ such that $\sup_t \mathbb{E}\,|w_t|\, 1(|w_t| > \lambda_\epsilon) < \epsilon$. Then define $w_{1,t} = w_t 1(|w_t| \leq \lambda_\epsilon)$ and $w_{2,t} = w_t - w_{1,t} = w_t 1(|w_t| > \lambda_\epsilon)$.

As in Jansson (2002, Proof of Lemma 5), we follow similar steps as in Hall and Heide (1980, Proof of Theorem 2.22). By the Marcinkiewicz-Zygmund inequality there exists a positive constant

C_1 not depending on n such that

$$\mathbb{E}\left|(nh)^{-1}\sum_{t=1}^{n-i} k_t k_{t+i}\sigma_t\sigma_{t+i} w_t\right| \leq C_1 (nh)^{-1}\left(\sum_{t=1}^{n-i} k_t^2(\tau)k_{t+i}^2(\tau)\sigma_t^2\sigma_{t+i}^2\,\mathbb{E}\,w_{t-j}^2\right)^{1/2}$$
$$\leq C_2\left[(nh)^{-1}\left(\sum_{t=1}^{n-i} k_t^2 k_{t+i}^2\sigma_t^2\sigma_{t+i}^2\,\mathbb{E}\,w_{1,t-j}^2\right)^{1/2} + (nh)^{-1}\left(\sum_{t=1}^{n-i} k_t^2 k_{t+i}^2\sigma_t^2\sigma_{t+i}^2\,\mathbb{E}\,w_{2,t-j}^2\right)^{1/2}\right].$$

As $|w_{1,t-j}| \leq \lambda_\epsilon$ and $\sup_t \sigma_t < \infty$, the first part can be bounded by

$$\begin{aligned}
(nh)^{-1}\left(\sum_{t=1}^{n-i} k_t^2(\tau)k_{t+i}^2(\tau)\sigma_t^2\sigma_{t+i}^2\,\mathbb{E}\,w_{1,t-j}^2\right)^{1/2} &\leq (nh)^{-1}\left(\sup_{s\in[0,1]}\sigma(s)^4\sum_{t=1}^{n-i} k_t^2 k_{t+i}^2\,\mathbb{E}\,w_{1,t-j}^2\right)^{1/2}\\
&\leq (nh)^{-1}C_1\left(\lambda_\epsilon^2\sum_{t=1}^{n-i} k_t^2(\tau)k_{t+i}^2(\tau)\right)^{1/2}\\
&\leq (nh)^{-1}C_1\left(\lambda_\epsilon^2\sum_{t=1\vee n(\tau-h))}^{n\wedge n(\tau+h)} k_t^2(\tau)k_{t+i}^2(\tau)\right)^{1/2}\\
&\leq (nh)^{-1}C_1\lambda_\epsilon\left(2nh\sup_s\left\{K(s)^4\right\}\right)^{1/2} = (nh)^{-1/2}C_2\lambda_\epsilon,
\end{aligned}$$

where for the last two inequalities we use Lemma A.1.

Using that $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$ for any n-dimensional vector $\mathbf{x}$ and $\mathbb{E}\,|w_{2,t}| \leq \epsilon$, we get for the second part

$$\begin{aligned}
(nh)^{-1}\left(\sum_{t=1}^{n-i} k_t^2(\tau)k_{t+i}^2(\tau)\sigma_t^2\sigma_{t+i}^2\,\mathbb{E}\,w_{2,t-j}^2\right)^{1/2} &\leq (nh)^{-1}\sup_{s\in[0,1]}\left\{\sigma(s)^2\right\}\sum_{t=1}^{n-i} k_t k_{t+i}\,\mathbb{E}\,|w_{2,t-j}|\\
&\leq (nh)^{-1}\sup_{s\in[0,1]}\left\{\sigma(s)^2\right\}\sum_{t=1\vee n(\tau-h))}^{n\wedge n(\tau+h)} k_t k_{t+i} \leq C_3\epsilon,
\end{aligned}$$

where the last line follows as in the previous part.

Concluding, we find that $\phi_n \leq C[(nh)^{-1/2}\lambda_\epsilon + \epsilon] = o(1) + C\epsilon$. As we can make ϵ arbitrarily small, it follows that $\phi_n \to 0$ as $n \to \infty$.

Next we look at η_n. First note that $\sum_{j=0}^{\infty}\sum_{m=0}^{\infty}|\psi_j|\,|\psi_m| = \left(\sum_{j=0}^{\infty}|\psi_j|\right)^2 < \infty$. Next, note that by Jensen's inequality*

$$\begin{aligned}
&\mathbb{E}\left|(nh)^{-1/2}\mathbb{1}_{m\neq j+i}\sum_{t=1}^{n-i} k_t(\tau)k_{t+i}(\tau)\sigma_t\sigma_{t+i}\varepsilon_{t-j}\varepsilon_{t+i-m}\right|\\
&\leq (nh)^{-1/2}\mathbb{1}_{m\neq j+i}\left[\mathbb{E}\left(\sum_{t=1}^{n-i} k_t(\tau)k_{t+i}(\tau)\sigma_t\sigma_{t+i}\varepsilon_{t-j}\varepsilon_{t+i-m}\right)^2\right]^{1/2}\\
&= (nh)^{-1/2}\mathbb{1}_{m\neq j+i}\left[\sum_{s=1}^{n-i}\sum_{t=1}^{n-i} k_s(\tau)k_t(\tau)k_{s+i}(\tau)k_{t+i}(\tau)\sigma_s\sigma_t\sigma_{s+i}\sigma_{t+i}\,\mathbb{E}\,\varepsilon_{s-j}\varepsilon_{t-j}\varepsilon_{s+i-m}\varepsilon_{t+i-m}\right]^{1/2}.
\end{aligned}$$

*$\mathbb{E}\,|X| = \mathbb{E}\sqrt{X^2} \leq \sqrt{\mathbb{E}\,X^2}$.

Note that $\mathbb{1}_{m\neq j+i}\,\mathbb{E}\,\varepsilon_{s-j}\varepsilon_{t-j}\varepsilon_{s+i-m}\varepsilon_{t+i-m}$ can only be non-zero if $t = s$.[†] Therefore we can deduce that

$$\begin{aligned}
&\mathbb{E}\left|(nh)^{-1/2}\mathbb{1}_{m\neq j+i}\sum_{t=1}^{n-i}k_t(\tau)k_{t+i}(\tau)\sigma_t\sigma_{t+i}\varepsilon_{t-j}\varepsilon_{t+i-m}\right| \\
&\leq (nh)^{-1/2}\left[\sum_{t=1}^{n-i}k_t^2(\tau)k_{t+i}^2(\tau)\sigma_t^2\sigma_{t+i}^2\right]^{1/2} \\
&\leq (nh)^{-1/2}\left[2nh\sup_{s\in[0,1]}\left\{\sigma(s)^4\right\}\sup_s\left\{K(s)^4\right\}\right]^{1/2}\leq C,
\end{aligned}$$

where the last line follows again from the properties of the kernel function as before. This shows that $\limsup_{n\to\infty}\eta_n<\infty$ which completes the proof. □

Proof of Lemma A.8. We first prove statement (i). For this, we look at the quantity $\mathbb{E}\left(N_{\tau_0,n}(\tau)\right)$ more closely.

$$\begin{aligned}
\mathbb{E}\left(N_{\tau_0,n}(\tau)\right) &= \mathbb{E}\left((nh)^{1/2}\left(\hat{m}(\tau_0+\tau h)-m(\tau_0+\tau h)\right)\right) \\
&= (nh)^{1/2}\,\mathbb{E}\left((nh)^{-1}\sum_{t=1}^{n}k_t(\tau_0+\tau h)y_t-m(\tau_0+\tau h)\right) \\
&= (nh)^{1/2}h^{-1}\left(n^{-1}\sum_{t=1}^{n}k_t(\tau_0+\tau h)m(t/n)-m(\tau_0+\tau h)\right)
\end{aligned}$$

We can now apply the integral approximation bound B.1 used in the proof of Lemma A.4 with $g(x) = K\left(\frac{\tau_0+\tau h-x}{h}\right)m(x)$. We know that $g(x)$ is continuous and Riemann-integrable. We obtain the following bound, where the arguments in the integral expression follow from setting $\omega = \frac{\tau_0+\tau h-t/n}{h}$

$$\left|n^{-1}\sum_{t=1}^{n}k_t(\tau_0+\tau h)m(t/n)-\int_0^1K(\omega)m(\tau_0+\tau h-\omega h)d\omega\right|\leq w_g(n^{-1}), \tag{B.2}$$

where

$$\begin{aligned}
w_g(n^{-1}) &= \sup_{|x_1-x_2|\leq 1/n}\left|K\left(\frac{\tau_0+\tau h-x_1}{h}\right)m(x_1)-K\left(\frac{\tau_0+\tau h-x_2}{h}\right)m(x_2)\right| \\
&= \sup_{|x_1-x_2|\leq 1/n}\left|m(x_1)\left(K\left(\frac{\tau_0+\tau h-x_1}{h}\right)-K\left(\frac{\tau_0+\tau h-x_2}{h}\right)\right)\right. \\
&\qquad \left.-K\left(\frac{\tau_0+\tau h-x_2}{h}\right)(m(x_2)-m(x_1))\right| \\
&= \sup_{|x_1-x_2|\leq 1/n}\left|C_1\frac{|x_1-x_2|}{h}-C_2|x_1-x_2|\right|\leq C_2(nh)^{-1}-C_2n^{-1}=O\left((nh)^{-1}\right)
\end{aligned}$$

[†]The case $m = j+i$ is ruled out by the indicator function, while for the other option it would be required that both $s-j = t+i-m$ and $t-j = s+i-m$, which is impossible.

Additionally, we can show that

$$\sup_{\tau\in[-1,1]}\left|(nh)^{1/2}\left(\int_0^1 K(\omega)m(\tau_0+\tau h-\omega h)d\omega - m(\tau_0+\tau h)\right) - B_{as}(\tau_0)\right| = o(1) \tag{B.3}$$

A second order Taylor approximation of $m(\cdot)$ around $\tau_0+\tau h$ yields:

$$m(\tau_0+\tau h-\omega h)\approx m(\tau_0+\tau h)+m'(\tau_0+\tau h)(-\omega h)+m''(\tau_0+\tau h)\frac{(\omega h)^2}{2}$$

Replacing $m(\tau_0+\tau h-\omega h)$ in the above integral by its approximation, the following three parts arise

$$\begin{aligned}
(1) &= \int_0^1 K(\omega)m(\tau_0+\tau h)d\omega = m(\tau_0+\tau h)\int_0^1 K(\omega)d\omega = m(\tau_0+\tau h)\\
(2) &= \int_0^1 K(\omega)m'(\tau_0+\tau h)(-\omega h)d\omega = -m'(\tau_0+\tau h)h\int_0^1 K(\omega)d\omega = 0\\
(3) &= \int_0^1 K(\omega)m''(\tau_0+\tau h)\frac{(\omega h)^2}{2}d\omega = m''(\tau_0+\tau h)h^2\int_0^1 K(\omega)\frac{\omega^2}{2}d\omega
\end{aligned}$$

Combining this with the left hand side of (B.3), we are able to show its asymptotic negligibility. Specifically, part (1) cancels with $m(\tau_0+\tau h)$, part (2) equals zero and the last part cancels with the asymptotic bias expression with the help of the Lipschitz continuity of $m''(\cdot)$.

$$\begin{aligned}
&\sup_{\tau\in[-1,1]}\left|(nh)^{1/2}\left(\int_0^1 K(\omega)m(\tau_0+\tau h-\omega h)d\omega - m(\tau_0+\tau h)\right) - B_{as}(\tau_0)\right|\\
&\approx \sup_{\tau\in[-1,1]}\left|(nh)^{1/2}\left((1)+(2)+(3)-m(\tau_0+\tau h)\right)-B_{as}(\tau_0)\right|\\
&= \sup_{\tau\in[-1,1]}\left|(nh)^{1/2}h^2\int_0^1 K(\omega)\frac{\omega^2}{2}d\omega\left(m''(\tau_0+\tau h)-m''(\tau_0)\right)\right|\\
&\le \sup_{\tau\in[-1,1]}\left|C_1(nh)^{1/2}h^2\tau h\int_0^1 K(\omega)\frac{\omega^2}{2}d\omega\right|\\
&\le C_2h^{7/2}n^{1/2}\le C_3n^{-1/5}=o(1),
\end{aligned}$$

where we use the fact that $h=Cn^{-1/5}$ to get to the final conclusion. Hence, (B.3) holds and we can use it together with (B.2) in the following way

$$\begin{aligned}
&\sup_{\tau\in[-1,1]}\left|\mathbb{E}\left(N_{\tau_0,n}(\tau)\right)-B_{as}(\tau_0)\right|\\
&= \sup_{\tau\in[-1,1]}\Bigg|(nh)^{1/2}h^{-1}\left(n^{-1}\sum_{t=1}^n k_t(\tau_0+\tau h)m(t/n)-m(\tau_0+\tau h)\right)-B_{as}(\tau_0)\\
&\quad-\left((nh)^{-1/2}\left(\int_0^1 K(\omega)m(\tau_0+\tau h-\omega h)d\omega-m(\tau_0+\tau h)\right)-B_{as}(\tau_0)\right)\Bigg|\\
&\le C(nh)^{1/2}h^{-1}(nh)^{-1}=C(nh^3)^{-1/2}=o(1).
\end{aligned}$$

Thus, the proof of statement (i) is complete.

We deal with statement (ii) next. Since both terms stated in $N^*_{\tau_0,n}$ are random, it is more convenient to work with centered versions and to divide the left-hand side of the statement into two parts, a random and a deterministic part. Let

$$\begin{aligned}
R_1(\tau) &:= \mathbb{E}^*\left(\hat{m}^*(\tau_0+\tau h)\right) - \mathbb{E}\left(\mathbb{E}^*\left(\hat{m}^*(\tau_0+\tau h)\right)\right) \\
R_2(\tau) &:= \tilde{m}(\tau_0+\tau h) - \mathbb{E}\left(\tilde{m}(\tau_0+\tau h)\right) \\
D_1(\tau) &:= \mathbb{E}\left(\mathbb{E}^*\left(\hat{m}^*(\tau_0+\tau h)\right)\right) \\
D_2(\tau) &:= \mathbb{E}\left(\tilde{m}(\tau_0+\tau h)\right)
\end{aligned}$$

and we get that

$$\mathbb{E}^*\left(N^*_{\tau_0,n}(\tau)\right) - B_{as}(\tau_0) = (nh)^{1/2}\left(R_1(\tau)-R_2(\tau)\right) + (nh)^{1/2}\left(D_1(\tau)-D_2(\tau)\right) - B_{as}(\tau_0)$$

Consider the deterministic parts first. For D_1, we will need to consider one sum at a time

$$D_1(\tau) = \mathbb{E}\left(\mathbb{E}^*\left(\hat{m}^*(\tau_0+\tau h)\right)\right) = (nh\tilde{h})^{-1}\sum_{t=1}^{n} k_t(\tau_0+\tau h)\,\underbrace{n^{-1}\sum_{i=1}^{n}\tilde{k}_t(i/n)m(i/n)}_{D_1^I(\tau)}$$

For both sums, we will use the integral approximation bound. For $D_1^I(\tau)$, we can use it in a similar way as in (B.2)

$$\left|n^{-1}\sum_{i=1}^{n}\tilde{k}_t(i/n)m(i/n) - \int_0^1 K(\omega)m(t/n-\omega\tilde{h})d\omega\right| = O\left((n\tilde{h})^{-1}\right) \tag{B.4}$$

We will also need the following integral approximation bound

$$\begin{aligned}
&\Big|n^{-1}\sum_{t=1}^{n} k_t(\tau_0+\tau h)\int_0^1 K(\omega)m(t/n-\omega\tilde{h})d\omega \\
&\qquad\qquad - \int_0^1 K(\eta)\int_0^1 K(\omega)m(\tau_0+\tau h-\eta h-\omega\tilde{h})d\omega d\eta\Big| \\
&\leq \sup_{|x_1-x_2|<n^{-1}}\Big|K\left(\frac{\tau_0+\tau h-x_1}{h}\right)\int_0^1 K(\omega)m(x_1-\omega\tilde{h})d\omega \\
&\qquad\qquad - K\left(\frac{\tau_0+\tau h-x_2}{h}\right)\int_0^1 K(\omega)m(x_2-\omega\tilde{h})d\omega\Big| \\
&\leq (nh)^{-1}
\end{aligned} \tag{B.5}$$

Subsequently, for D_2 we will use:

$$D_2(\tau) = \mathbb{E}\left(\tilde{m}(\tau_0+\tau h)\right) = (n\tilde{h})^{-1}\sum_{t=1}^{n}\tilde{k}_t(\tau_0+\tau h)m(t/n)$$

Again, the integral approximation bound gives us

$$\left|(n\tilde{h})^{-1}\sum_{t=1}^{n}\tilde{k}_t(\tau_0+\tau h)m(t/n)-\int_0^1 K(\omega)m(\tau_0+\tau h-\omega\tilde{h})d\omega\right| = O\left(n^{-1}\tilde{h}^{-2}\right) \tag{B.6}$$

Combining the results above yields an expression for $D_1(\tau)-D_2(\tau)$ with O-terms which are uniform in τ due to the Lipschitz properties of $m(\cdot)$ and $K(\cdot)$ as well as the boundedness of τ. To see this, we will add and subtract the corresponding parts to be able to use the three integral approximations (B.4), (B.5) and (B.6).

$$\begin{aligned}
&D_1(\tau)-D_2(\tau)\\
&=(nh\tilde{h})^{-1}\sum_{t=1}^{n}k_t(\tau_0+\tau h)n^{-1}\sum_{i=1}^{n}\tilde{k}_t(i/n)m(i/n)-(n\tilde{h})^{-1}\sum_{t=1}^{n}\tilde{k}_t(\tau_0+\tau h)m(t/n)\\
&=(nh\tilde{h})^{-1}\sum_{t=1}^{n}k_t(\tau_0+\tau h)\left[n^{-1}\sum_{i=1}^{n}\tilde{k}_t(i/n)m(i/n)-\int_0^1 K(\omega)m(t/n-\omega\tilde{h})d\omega\right]\\
&\quad+(nh\tilde{h})^{-1}\sum_{t=1}^{n}k_t(\tau_0+\tau h)\int_0^1 K(\omega)m(t/n-\omega\tilde{h})d\omega-\int_0^1 K(\omega)m(\tau_0+\tau h-\omega\tilde{h})d\omega+O\left(n^{-1}\tilde{h}^{-2}\right)\\
&=(nh\tilde{h})^{-1}\sum_{t=1}^{n}k_t(\tau_0+\tau h)O\left((n\tilde{h})^{-1}\right)+O\left(n^{-1}\tilde{h}^{-2}\right)\\
&\quad+(nh\tilde{h})^{-1}\sum_{t=1}^{n}k_t(\tau_0+\tau h)\int_0^1 K(\omega)m(t/n-\omega\tilde{h})d\omega-\int_0^1 K(\eta)\int_0^1 K(\omega)m(\tau_0+\tau h-\eta h-\omega\tilde{h})d\omega d\eta\\
&\quad+\int_0^1 K(\eta)\int_0^1 K(\omega)m(\tau_0+\tau h-\eta h-\omega\tilde{h})d\omega d\eta-\int_0^1 K(\omega)m(\tau_0+\tau h-\omega\tilde{h})d\omega\\
&=\int_0^1 K(\eta)\int_0^1 K(\omega)m(\tau_0+\tau h-\eta h-\omega\tilde{h})d\omega d\eta-\int_0^1 K(\omega)m(\tau_0+\tau h-\omega\tilde{h})d\omega\\
&\quad+O\left(n^{-1}\tilde{h}^{-2}h^{-1}\right)+O\left(n^{-1}\tilde{h}^{-2}\right)+O\left(n^{-1}\tilde{h}^{-1}h^{-2}\right)\\
&=\int_0^1 K(\eta)\int_0^1 K(\omega)\left[m(\tau_0+\tau h-\eta h-\omega\tilde{h})-m(\tau_0+\tau h-\omega\tilde{h})\right]d\omega d\eta+O\left(n^{-1}\tilde{h}^{-1}h^{-2}\right)
\end{aligned}$$

To finally show that

$$\sup_{\tau\in[-1,1]}\left|(nh)^{1/2}(D_1(\tau)-D_2(\tau))-B_{as}(\tau_0)\right| = o(1) \tag{B.7}$$

we can switch the order of integration and apply a second order Taylor approximation to $m(\tau_0+\tau h-\eta h-\omega\tilde{h})$ around the point $\tau_0+\tau h-\omega\tilde{h}$ to replace the term in square brackets by

$$m(\tau_0+\tau h-\eta h-\omega\tilde{h})-m(\tau_0+\tau h-\omega\tilde{h})\approx m'(\tau_0+\tau h-\omega\tilde{h})(-\eta h)+m''(\tau_0+\tau h-\omega\tilde{h})\frac{(\eta h)^2}{2}$$

The first part contains the expectation of the kernel function and thus will be equal to zero. Hence, we get

$$D_1(\tau)-D_2(\tau)\approx\int_0^1 K(\omega)m''(\tau_0+\tau h-\omega\tilde{h})h^2\int_0^1\frac{K(\eta)\eta^2}{2}d\eta d\omega+O\left(n^{-1}\tilde{h}^{-1}h^{-2}\right).$$

To continue from here, we can use the Lipschitz property of $m''(\cdot)$ to see that

$$m''(\tau_0 + \tau h - \omega\tilde{h}) \leq C|\tau h - \omega\tilde{h}| + m''(\tau_0)$$

and replace $m''(\tau_0 + \tau h - \omega\tilde{h})$ by this bound

$$\begin{aligned} D_1(\tau) - D_2(\tau) &\leq B_{as}(\tau_0) + C\tau h^3 \int_0^1 K(\omega)d\omega \int_0^1 \frac{K(\eta)\eta^2}{2} d\eta + O\left(n^{-1}\tilde{h}^{-1}h^{-2}\right) \\ &= B_{as}(\tau_0) + O\left(h^3\right) + O\left(n^{-1}\tilde{h}^{-1}h^{-2}\right). \end{aligned}$$

This is sufficient to show that (B.7) holds.

Next, we look at the random part $R_1(\tau) - R_2(\tau)$. First, consider $R_1(\cdot)$. It is easy to see that

$$\begin{aligned} R_1(\tau) &= \mathbb{E}^* \left(\hat{m}^*(\tau_0 + \tau h)\right) - \mathbb{E}\left(\mathbb{E}^* \left(\hat{m}^*(\tau_0 + \tau h)\right)\right) \\ &= \mathbb{E}^* \left((nh)^{-1} \sum_{t=1}^n k_t(\tau_0 + \tau h)\tilde{m}(t/n)\right) + \mathbb{E}^* \left((nh)^{-1} \sum_{t=1}^n k_t(\tau_0 + \tau h) z_t^*\right) - D_1(\tau) \\ &= (nh\tilde{h})^{-1} \sum_{t=1}^n k_t(\tau_0 + \tau h) n^{-1} \sum_{i=1}^n \tilde{k}_t(i/n) y_i - D_1(\tau) \\ &= (nh\tilde{h})^{-1} \sum_{t=1}^n k_t(\tau_0 + \tau h) n^{-1} \sum_{i=1}^n \tilde{k}_t(i/n) \left(m(i/n) + z_i\right) - D_1(\tau) \\ &= (nh\tilde{h})^{-1} \sum_{t=1}^n k_t(\tau_0 + \tau h) n^{-1} \sum_{i=1}^n \tilde{k}_t(i/n) z_i \end{aligned}$$

so that for fluctuations in $R_1(\cdot)$ we get

$$R_1(\tau) - R_1(\vartheta) = n^{-2}(h\tilde{h})^{-1} \sum_{t=1}^n \sum_{i=1}^n \left(k_t(\tau_0 + \tau h) - k_t(\tau_0 + \vartheta h)\right) \tilde{k}_t(i/n) z_i$$

so that with the help of the Lipschitz property of $K(\cdot)$ and Lemmas A.2 and A.1 we obtain for the expected squared fluctuations

$$\begin{aligned} &\mathbb{E}\left|R_1(\tau) - R_1(\vartheta)\right|^2 \\ &\quad\leq n^{-4}(h\tilde{h})^{-2} \sum_{t,s,i=1}^n \sum_{k=-n+1}^{n-1} \Big|\left(k_t(\tau_0 + \tau h) - k_t(\tau_0 + \vartheta h)\right) \\ &\qquad\qquad \times \left(k_s(\tau_0 + \tau h) - k_s(\tau_0 + \vartheta h)\right) \tilde{k}_t(i/n)\tilde{k}_s((i-k)/n)\Big| \left|\sigma_i \sigma_{i-k}\right| \left|R(k)\right| \\ &\quad\leq n^{-4}(h\tilde{h})^{-2} \sum_{t,s,i=1}^n C_1|\tau - \vartheta|^2 \tilde{k}_t(i/n) \sup_k \left\{\tilde{k}_s((i-k)/n)\right\} \left[\sigma(\tau)^2 + C_2 h\right] \sum_{k=-n+1}^{n-1} |R(k)| \\ &\quad\leq C_3 n^{-4}(h\tilde{h})^{-2}(nh)^2 n\tilde{h}|\tau - \vartheta|^2 \sup_s \left\{K(s)^2\right\}, \end{aligned}$$

such that

$$n\tilde{h}\,\mathbb{E}\left|R_1(\tau) - R_1(\vartheta)\right|^2 \leq C|\tau - \vartheta|^2. \tag{B.8}$$

Second, we consider $R_2(\cdot)$. Similarly, we get

$$
\begin{aligned}
&\mathbb{E}\,|R_2(\tau) - R_2(\vartheta)|^2 \\
&\quad \le (n\tilde{h})^{-2} \sum_{t=1}^{n} \sum_{k=-n+1}^{n-1} |(k_t(\tau_0 + \tau h) - k_t(\tau_0 + \vartheta h))\,(k_{t-k}(\tau_0 + \tau h) - k_{t-k}(\tau_0 + \vartheta h))|\,|\sigma_t \sigma_{t-k}|\,|R(k)|
\end{aligned}
$$

and as for $R_1(\cdot)$ it holds that

$$
n\tilde{h}\,\mathbb{E}\,|R_2(\tau) - R_2(\vartheta)|^2 \le C|\tau - \vartheta|^2. \tag{B.9}
$$

Letting $R(\tau) := R_1(\tau) - R_2(\tau)$ and combining both (B.8) and (B.9) using the triangle inequality leaves us with

$$
\begin{aligned}
n\tilde{h}\,\mathbb{E}\,|R(\tau) - R(\vartheta)|^2 &= n\tilde{h}\,\mathbb{E}\,|R_1(\tau) - R_1(\vartheta) - (R_2(\tau) - R_2(\vartheta))|^2 \\
&\le n\tilde{h}\,\mathbb{E}\,|R_1(\tau) - R_1(\vartheta)|^2 + n\tilde{h}\,\mathbb{E}\,|R_2(\tau) - R_2(\vartheta)|^2 \\
&\qquad + 2n\tilde{h}\,\mathbb{E}\,|(R_1(\tau) - R_1(\vartheta))\,(R_2(\tau) - R_2(\vartheta))| \\
&\le C|\tau - \vartheta|^2.
\end{aligned} \tag{B.10}
$$

This is exactly what we need to ensure that all conditions of Theorem 12.3 in Billingsley (1968) are met to obtain tightness of $R(\cdot)$. The condition of tightness of $R(0)$ is satisfied in this case, because for $\tau = 0$ we only look at $R(\tau_0)$ which is tight since pointwise weak convergence has been established. Furthermore, we verified the moment condition for $\gamma = \alpha = 2$ and the function $F(\cdot)$ being the identity. It is known that tightness of $R(\cdot)$ implies stochastic equicontinuity (see e.g. Newey (1991)) and we get that for $\eta > 0$ and $\kappa > 0$, there exists $\lambda > 0$ and $n_0 \in \mathbb{N}$ such that

$$
\mathbb{P}\left(\sup_{|\tau - \vartheta| \le \lambda} (n\tilde{h})^{1/2}\,|R(\tau) - R(\vartheta)| > \kappa \right) < \eta \quad \forall\, n \ge n_0
$$

and therefore,

$$
\lim_{\lambda \searrow 0} \limsup_{n\to\infty} \mathbb{P}\left(\sup_{|\tau - \vartheta| \le \gamma} (n\tilde{h})^{1/2}\,|R(\tau) - R(\vartheta)| > \kappa \right) < \eta \quad \forall\, n \ge n_0
$$

We need stochastic equicontinuity of $R(\cdot)$ to get to the desired result that

$$
\sup_{\tau \in [-1,1]} (nh)^{1/2} |R(\tau)| = o_p(1) \tag{B.11}
$$

which together with (B.7) is sufficient to show that the statement holds. Stochastic equicontinuity allows us to discretize the interval $[-1, 1]$ over which we take the supremum. It allows us to look at discrete points in order to obtain the O_p- and o_p-statements which are uniform in $\tau \in [-1, 1]$. First, to show the validity of (B.11), we follow analogous steps as in the derivation of (B.10) to obtain the three bounds which are uniform for $\tau \in [-1,1]$

$$n\tilde{h}\,\mathbb{E}\,|R_1(\tau)|^2 \leq C_1$$
$$n\tilde{h}\,\mathbb{E}\,|R_2(\tau)|^2 \leq C_2$$
$$(n\tilde{h})^{1/2}\,\mathbb{E}\,|R(\tau)|^2 \leq C_3,$$

where we use the first two bounds to derive a squared version of the third in the same way as for the fluctuation case above. Using Chebychev's inequality we get

$$(n\tilde{h})^{1/2}\mathbb{P}\left(|R(\tau)| \geq \kappa\right) \leq (n\tilde{h})^{1/2}\frac{\mathbb{E}\,|R(\tau)|}{\kappa} \leq C$$

which implies

$$\sup_{\tau \in [-1,1]} (n\tilde{h})^{1/2}|R(\tau)| = O_p(1)$$

This, together with the fact that $h = o(\tilde{h})$, allows us to conclude that (B.11) holds. Hence, after rewriting $\mathbb{E}^*\left(N^*_{\tau_0,n}(\tau)\right) - B_{as}(\tau_0)$ in terms of the quantities $R_1(\tau)$, $R_2(\tau)$, $D_1(\tau)$ and $D_2(\tau)$ and considering the deterministic and random part in turn, we showed that statement (ii) of the lemma holds. □

References

Billingsley, P. (1968). Convergence of probability measures. Wiley, New York.

Bühlmann, P. (1998). Sieve bootstrap for smoothing in nonstationary time series. Annals of Statistics 26, 48-83.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. Journal of Econometrics 136, 163-188.

Chu, C.-K. and J.S. Marron (1991). Comparison of two bandwidths selectors with dependent errors. The Annals of Statistics 19(4), 1906-1918.

Davidson, J. (2002). Stochastic Limit Theory (2nd edition). Oxford: Oxford University Press.

Davidson, R. and E. Flachaire (2008). The wild bootstrap, tamed at last. Journal of Econometrics 146, 162-169.

Fan, J. (1992). Design-adaptive nonparametric regression. Journal of the American Statistical Association 87(420), 998-1004.

Franco, B., Bader W., Toon G.C., Bray C., Perrin A., Fischer E.V., Sudo K., Boone C.D., Bovya B., Lejeune B., Servais C. and E. Mahieu (2015). Retrieval of ethane from ground-based FTIR solar spectra using improved spectroscopy: Recent burden increase above Jungfraujoch. Journal of Quantitative Spectroscopy and Radiative Transfer 160, 36-49.

Franco, B., Mahieu, E., Emmons, L.K., Tzompa-Sosa, Z.A., Fischer, E.V., Sudo, K., Bovy, B., Conway, S., Griffin, D., Hannigan, J.W., Strong, K. and K.A. Walker (forthcoming). Evaluating ethane and methane emissions associated with the development of oil and natural gas extraction in North America. Environmental Research Letters, 11(4), 044010, doi:10.1088/1748-9326/11/4/044010, 2016.

Franses, P.H. and T.J. Vogelsang (2005). Are winters getting warmer? Environmental Modeling and Software 20, 1449-1455.

Gasser, T. and H.G. Müller (1984). Estimating regression functions and their derivatives by the kernel method. Scandinavian Journal of Statistics 11, 171-185.

Gonzalés, A. and T. Teräsvirta (2008). Modelling autoregressive processes with a shifting mean. Studies in Nonlinear Dynamics & Econometrics 12(1).

Hall, P. and C.C. Heyde (1980). Martingale Limit Theory and Its Application. New York: Springer Academic Press.

Jansson, M. (2002). Consistent covariance matrix estimation for linear processes. Econometric Theory 18, 1449-1459.

Leucht, A. and M. H. Neumann (2013). Dependent wild bootstrap for degenerate U- and V-statistics. Journal of Multivariate Analysis 117, 257-280.

McKitrick, R. and T.J. Vogelsang (2014). HAC robust trend comparisons among climate series with possible level shifts. Environmetrics 25(7), 528-547.

Nadaraya, E.A. (1964). On estimating regression. Theory Prob. Appl. 9, 141-142.

Neumann, M.H. (1998). Pointwise confidence intervals in nonparametric regression with heteroskedastic error structure. Statistics 29(1), 1-36.

Neumann, M.H. and J. Polzehl (1998). Simultaneous bootstrap confidence bands in nonparametric regression. Journal of Nonparametric Statistics 9, 307-333.

Newey, W. (1991). Uniform convergence in probability and stochastic equicontinuity. Econometrica 59(4), 1161-1167.

Shao, X. (2010). The dependent wild bootstrap. Journal of the American Statistical Association 105, 218-235.

Simpson, I.J., Sulbaek Andersen, M.P., Meinardi, S., Bruhwiler, L., Blake, N.J., Helmig, D., Rowland, F. S. and D. R, Blake (2012). Longterm decline of global atmospheric ethane concentrations and implications for methane Nature 488, 490-494.

Smeekes, S. and J.-P. Urbain (2014). A multivariate invariance principle for modified wild bootstrap methods with an application to unit root testing. Maastricht University Research Papers, RM/14/008.

Vinciguerra, T., Yao, S., Dadzie, J., Chittams, A., Deskins, T., Ehrman, S. and R.R. Dickerson (2015). Regional air quality impacts of hydraulic fracturing and shale natural gas activity: evidence from ambient VOC observations. Atmos. Environ. 110, 144-50.

Watson, G.S. (1964). Smooth regression analysis. Sankhyā Ser. A 26, 359-372.

Wu, W.B. and Z. Zhao (2007). Inference of trends in time series. Journal of the Royal Statistical Society Series B 69, 391-410.

Yurganov, L.N., Blumenstock, T., Grechko, E.I., Hase, F., Hyer, E.J., Kasischke, I.S., Koike, M., Kondo, Y., Kramer, I., Leung, F.-Y., Mahieu, E., Mellqvist, J., Notholt, J., Novelli, P.C., Rinsland, C.P., Scheel, H.E., Schulz, A., Strandberg, A., Sussmann, R., Tanimoto, H., Velazco, V., Zander, R. and Y. Zhao (2004). A quantitative assessment of the 1998 carbon monoxide emission anomaly in the Northern Hemisphere based on total column and surface concentration measurements. Journal of Geophysical Research 109, D15305.

Yurganov, L. N., Duchatelet, P., Dzhola, A. V., Edwards, D. P., Hase, F., Kramer, I., Mahieu, E., Mellqvist, J., Notholt, J., Novelli, P.C., Rockmann, A., Scheel, H.E., Schneider, M., Schulz, A., Strandberg, A., Sussmann, R., Tanimoto, H., Velazco, V., Drummond, J.R. and J.C. Gille (2005).

Increased Northern Hemispheric carbon monoxide burden in the troposphere in 2002 and 2003 detected from the ground and from space. Atmospheric Chemistry and Physics, 5(2), 563-573.